Semantically congruent visual stimuli in the improvement of auditory memory in older adults

Petra Fagerlund

School of Science

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 8.9.2017

Thesis supervisor:

Prof. Mikko Sams

Thesis advisor:

Ph.D., Docent Kaisa Tiippana

Aalto University
School of Science

AALTO UNIVERSITY
SCHOOL OF SCIENCE

ABSTRACT OF THE
MASTER'S THESIS

Author: Petra Fagerlund

Title: Semantically congruent visual stimuli in the improvement of auditory memory in older adults

Date: 8.9.2017 Language: English Number of pages: 6+73

Department of Neuroscience and Biomedical Engineering

Professorship: Human Neuroscience and Technology

Supervisor: Prof. Mikko Sams

Advisor: Ph.D., Docent Kaisa Tiippana

Our society is facing great structural changes as the proportion of elder population keeps ageing faster and faster. It is now increasingly important to find ways to keep the elderly population physically, socially and, especially, cognitively functional. However, more research is needed on normal cognitive ageing in natural-like environments. Since we live in a multisensory world, it is highly important to understand how ageing affects multisensory processing. From previous research it is known that young adults perform better in audiovisual conditions compared to unisensory conditions. Thus, this study investigates whether older adults can benefit from audiovisual stimuli in the similar way than young adults.

This study investigated the effects of audiovisual encoding on later auditory recognition memory performance in older adults (n=42, mean=71 years). The participants memorized auditory stimuli (sounds or spoken words) presented with semantically congruent visual stimuli (pictures or written words) during encoding. The participants performed six tests: sounds or spoken words presented with pictures, written words or alone. Audiovisual conditions were compared to the unisensory ones to examine whether older adults can benefit from visual cues in the encoding. Subsequent recognition memory performance was better for audiovisual conditions than for unisensory auditory conditions, and spoken words were remembered better than sounds. Additionally, mean reaction times were measured, and they were faster to unisensory condition of sounds than to audiovisual conditions of sounds. These results suggest that older adults can benefit from semantically congruent multisensory experiences by enhancing the encoding of both non-verbal and verbal materials, resulting in an improvement in their later recognition memory. The present study provides the first evidence that older adults can benefit from multisensory memory cues.

Keywords: recognition memory, audiovisual, semantic congruency, multisensory perception, multisensory memory, older adults

AALTO-YLIOPISTO
PERUSTIETEIDEN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Petra Fagerlund

Työn nimi: Semanttisesti kongruentit visuaaliset ärsykkeet auditiivisen muistin parantamisessa ikääntyvillä

Päivämäärä: 8.9.2017 Kieli: Englanti Sivumäärä: 6+73

Neurotieteen ja lääketieteellisen tekniikan laitos

Professuuri: Neurotieteet ja teknologia

Työn valvoja: Prof. Mikko Sams

Työn ohjaaja: Ph.D, dosentti Kaisa Tiippana

Yhteiskuntamme kohtaa suuria rakenteellisia muutoksia väestön ikääntyessä yhä nopeammin. Nyt on erityisen tärkeää löytää keinoja, joilla ikääntyvä väestö voidaan pitää etenkin kognitiivisesti toimintakykyisinä. Lisää tutkimusta kuitenkin kaivataan normaalista kognitiivisesta ikääntymisestä luonnonmukaisessa ympäristössä. Tässä moniaistisessa maailmassa on tärkeää ymmärtää, kuinka ikääntyminen vaikuttaa moniaistisen informaation käsittelyyn. Aikaisemman tutkimuksen pohjalta tiedetään, että nuoret aikuiset suoriutuvat paremmin audiovisuaalisissa tilanteissa yksiaistisiin tilanteisiin verrattuna. Näin ollen voidaan olettaa, että ikääntyneet hyötyvät audiovisuaalisista ärsykkeistä yhtä lailla.

Tässä tutkimuksessa tutkittiin audiovisuaalisen mieleenpainamisen vaikutuksia myöhempään tunnistusmuistisuoriutumiseen ikääntyvillä. Tutkittavat (n=42, ka=71 vuotta) painoivat muistiin auditiivisia ärsykkeitä (ääniä tai puhuttuja sanoja), jotka oli esitetty yhdessä semanttisesti kongruentin visuaalisen ärsykkeen kanssa (kuva tai kirjoitettu sana). Tutkittavat suorittivat kuusi koetta: ääni tai puhuttu sana esitettynä yhdessä kuvan tai kirjoitetun sanan kanssa, tai yksinään. Audiovisuaalisia tilanteita verrattiin yksiaistiseen. Myöhempi tunnistusmuistisuoriutuminen oli parempi audiovisuaalisissa tilanteissa kuin yksiaistisissa tilanteissa, ja puhutut sanat muistettiin ääniä paremmin. Myös reaktioajat mitattiin, ja ne olivat nopeampia äänen yksiaistiselle tilanteelle kuin äänen audiovisuaalisille tilanteille. Nämä tulokset osoittavat, että ikääntyvät voivat hyötyä semanttisesti kongruenteista moniaistillisista kokemuksista parantaen niin ei-verbaalisten kuin verbaalisten materiaalien mieleenpainamista, ja näin ollen parantaa ikääntyvien tunnistusmuistia. Tämä tutkimus osoittaa ensimmäisenä, että ikääntyvät voivat hyötyä moniaistisista muistivihjeistä.

Avainsanat: tunnistusmuisti, audiovisuaalinen, semanttinen kongruenssi, moniaistinen havaitseminen, moniaistinen muisti, normaali ikääntyminen

Preface

I want to thank my advisor Docent Kaisa Tiippana and her PhD student Jenni Heikkilä for their good and skilful guidance. Additionally I thank Professor Mikko Sams for being my supervisor and for evaluating my thesis. I also want to thank my mother, Marja Seppälä, for recruiting most of the participants for this study through her own connections. Thanks also belong to Eero Lähteenmäki since he gave me extra support and guidance during the writing process. I want to thank Paul Keighley for checking the grammar, and my dear friend Alisa Kopilow, who properly read my thesis and provided great extra help on the finalizing process of the thesis. At last, I want to thank all those older adults who participated in the study and made this research possible and successful.

We were successful with the study, and thus we are publishing an article about the findings in a scientific journal this year. The review process is already at a very advanced state. It got accepted in the publication, and the article can be expected to be published in the Multisensory Research journal during the Autumn 2017. The first author is Jenni Heikkilä, I am the second writer, and the third and the last author is my advisor Kaisa Tiippana. Jenni will include this study into her PhD thesis.

It has been very interesting to be a part of a real scientific research straight from the beginning until the very end. Co-operation with Kaisa and Jenni has been enjoyable and professional with a nice and warm touch. I have learned a lot about research, including recruiting volunteers, setting a study design, analysing data, and scientific writing. Although I am not now continuing my career in research, I am sure this experience has given me needed criticism and maturity for future challenges in the work life.

With few tears in my eyes and a melancholic feeling in my heart, I end this thesis and my six years of studies in Aalto University. I am very grateful and lucky for this education and growth I have been given in my university by our professors and by my dear fellow students. Need to admit still, I cannot wait what the future brings into my life. I am so ready for new adventures!

So for now, bye bye Aalto University and student life, and hello working life!

Otaniemi, 8.9.2017

Petra M. Fagerlund

Contents

Abstract ii

Abstract (in Finnish) iii

Preface iv

Contents v

1 Introduction 1

2 Background 4
2.1 Normal cognitive ageing 4
2.1.1 Processing speed slows down with ageing 5
2.1.2 Age-related changes in the frontal cortex weakens cognition . . 6
2.2 Memory functions in normal ageing 6
2.2.1 Human memory systems 7
2.2.2 Recognition memory 8
2.2.3 Effect of age on memory systems 9
2.3 Sensory decline in normal ageing 10
2.3.1 Hearing 10
2.3.2 Vision 11
2.4 The effect of age on multisensory processing 11
2.4.1 Non-semantic multisensory perception 12
2.4.2 Audiovisual speech perception 13
2.4.3 Supplementing speech understanding with visual text 17
2.5 Semantically congruent multisensory memory and congruency effect . 18
2.5.1 Continuous recognition tasks in evaluating semantic congruency effect 19
2.5.2 Two-phased recognition memory tests in evaluating semantic congruency effect 20
2.5.3 Current study 22

3 Research material and methods 24
3.1 Participants 24
3.2 Stimuli 24
3.3 Design and procedure 25
3.4 Data analysis 28
3.4.1 Mean reaction times 28
3.4.2 Signal detection theory 28
3.4.3 Analysing data with repeated measures ANOVA 30
3.4.4 Data analysis workflow 33

4 Results 34
4.1 d' showed enhanced memory performance in audiovisual conditions . 34
4.2 Hit rate showed different memory performance than d' 36
4.3 Criterion *c* varied significantly 37
4.4 Response times were longer for Sounds than for Spoken words 38

5 Discussion 39
5.1 Summary of multisensory findings 39
5.2 Sounds showed poorer performance compared to Spoken words 40
5.3 Why does semantic congruency help older adults? 41
5.4 Older adults seem to show a great picture-facilitation effect with verbal material 43
5.5 Explaining findings from the point of view of the CSTM model 45
5.6 How older adults compare to their younger counterparts 46
5.7 Older adults require more response time when processing complicated sounds that were originally presented audiovisually 47
5.8 Limitations 48
5.9 Future prospects 49

6 Conclusions 50

References 57

A Normality tests: CDF graphs and a table 58

B List of stimuli 63

C Participant consent forms 71

1 Introduction

The population is ageing rapidly and the ratio of the working age population to the elderly population is decreasing. The finnish elderly population has doubled over the past 45 years [1]. By the year 2060, it is projected that the proportion of the elderly population will grow to 29 % [2]. A similar trend is seen worldwide. There has been an increase of 48 % of people aged 60 years or over in 2015 compared to year 2000, and it has been projected that by 2050, the global population of older persons is expected to more than double its size from 2015 [3].

Alongside the increasing older population, the incidence of memory disorders is increasing. In Finland, memory disorders are already considered as a national chronic disease. It has been estimated that currently in Finland there are approximately 120,000 people suffering from mild or moderately severe memory disorders and an additional 120,000 people suffering from mild cognitive impairment [4]. Globally, today there are over 46 million people living with dementia and by 2050 this number is expected to be almost tripled [5]. In addition, the healthcare costs are estimated at US$ 604 billion per year at present and are set to increase even more quickly than the prevalence [5, 6]. Clearly these progressive memory disorders pose a global challenge for public health and national economy.

As a response to the rapid ageing of population, ageing, and its biological, behavioural and cognitive changes, have been a great interest among researchers for decades. In light of the evolving demographic changes of our society, one important future task for the research communities is to further our understanding of lifelong healthy ageing [7]. Ageing is a multi-factorial and multidimensional process involving physiological, psychological, and social alterations. Particularly, sensory impairments have an enormous impact on our lives and are closely related to intellectual functioning.

We experience our environment through multiple sensory systems, which ultimately ensure our everyday safety, quality of life, and social adjustment. Thus, it is interesting to understand how multisensory integration processing changes as a function of healthy ageing. Multisensory interactions, that is, sensory information from multiple senses, are ubiquitous in the nervous system and occur at early stage of perceptual processing. Perception has traditionally been viewed as a modular function with the different sensory modalities operating largely as separate and independent processes [8]. However, reports of multisensory interactions in various perceptual tasks and settings indicate that these interactions are the rule rather than the exception in human processing of sensory information [9, 10], and there exists a rapidly growing literature of the neuroanatomical, electrophysiological and neuroimaging studies that show that multisensory interactions can occur throughout processing [10–12].

Given the extensive changes of perceptual and cognitive processes and the underlying structural and functional brain changes during healthy ageing, it seems reasonable that multisensory integration performance is altered throughout the lifespan as well. Although we live in a multisensory world, older adult's memory has usually been studied concentrating on only one sensory modality at a time. Memory retrieval of auditory stimuli has gotten less attention than the memory retrieval of visual stimuli among researchers and partly the results are dissonant. Multisensory memory in older adults have, to date, received very little attention.

There is considerable evidence of multisensory processes in old age. In non-semantic audiovisual perception studies, simple stimuli e.g. light (visual) and a single-tone beep (auditive) are simultaneously presented together, and participants' reaction times on perceiving the stimuli are measured and examined. These studies have found that older adults obtain greater multisensory gains than young adults [13, 14], however, no age effect and slower reaction times for older adults have also been presented [15]. Audiovisual speech perception studies, where participants aim to recognize syllables, words or sentences articulated by a speaker in a video clip, provide another interesting view. These studies have shown that older adults lean more toward visual cues than their younger counterparts [16–19]. Additionally, it has been observed that signal content affects audiovisual enhancement because ageing undermines the ability to take advantage of visual speech information when the visual cues are degraded [20]. It has been also suggested that information with complex content can be better remembered if presented audiovisually with semantically congruent cues [21].

In multisensory memory studies, on the other hand, participants' ability to encode audiovisual stimuli, e.g. sounds (auditory) and pictures (visual), and how this influences later memory test, is investigated. These studies have revealed an audiovisual congruency effect, that is, a strong audiovisual integration, with young adults [22–29] and with children [30]. Although, contradictory results have been reported as well [31]. The congruency effect has been tested using semantic congruency, which refers to a condition where two different stimuli share the same semantic meaning, e.g. a picture of a cow and a sound of a cow mooing. In these studies, the stimulus trials have not necessarily been always semantically congruent but semantically incongruent, that is, e.g. a picture of a sheep but a sound of a cow mooing, or semantically meaningless, that is, one of the stimuli does not have any semantic meaning (e.g. white noise). The results then, have been drawn by comparing the performances in responding to semantically congruent stimuli, to semantically incongruent stimuli or to semantically meaningless stimuli, and the studies have been mainly conducted with young adults.

Multisensory memory research is basically divided into two trends on how to investigate congruency effect and semantic congruency. In continuous recognition tasks participants are responding already during the experiment, discriminating between items seen for the first or for the second time during a block of trials [22–25]. These studies have found that recognition of auditory stimuli benefits more from additional visual stimuli than other way around [23–25]. A two-part recognition task, on the other hand, consists of a study phase and a test phase. This study setting requires participants to hold the information in their memory for longer periods of time, which thus tests the memory itself rather than the recognition ability. These studies have shown that the later recognition memory performance is improved when semantically congruent audiovisual material is presented during the encoding in young adults [26–30] and in children [30].

Among multisensory memory studies, only one study has investigated older adults in addition to young adults. Luo et al. [32] applied a two-part recognition task to investigate which semantically congruent conditions enhanced the recognition memory performance of older adults. Contrary to the results gained with young adults (see e.g. [27, 28]), the older adults did not benefit from audiovisual condition where written words (visual) were accompanied with semantically congruent sounds (auditory). However, according to these findings, it is not known yet whether older adults can benefit from other audiovisual stimuli combinations.

The goal of the present study was therefore to investigate the effects of audiovisual encoding on later unisensory recognition memory performance in older adults. The exactly same paradigm with two-part recognition memory testing was used as Heikkilä et al. [28] used in their very recent study conducted with young adults, but in the current study the participants were older adults. The participants (n=42, mean=71 years) memorized auditory stimuli (sounds or spoken words) presented alone or together with a semantically congruent visual stimulus (pictures or written words) during memory encoding. The semantically congruent audiovisual stimuli pairs and unisensory stimuli were presented in separate blocks, which allowed the precision of recognition memory between congruent audiovisual and unisensory situations be compared by applying a sensitivity index *d'*. The mean reactions times were measured as well, although no statistically significant main effect was expected to be found. General hypothesis was that older adults can benefit from congruent audiovisual information in memory encoding in a similar way to young adults. Additionally, it was hypothesized that spoken words will be remembered better than sounds. The main research questions were: Do older adults benefit from semantically congruent audiovisual stimulus in the encoding phase? Can the auditory recognition memory be enhanced by semantically congruent visual stimuli presented during encoding?

2 Background

As life expectancy rises in industrialized countries, there is a growing need to understand the cognitive and neural declines that accompany normal ageing. Equally important is the understanding of multisensory perception and multisensory memory, since in everyday life, the majority of situations and events contain multisensory elements. This section sheds light on normal ageing in different fields; in cognition (Ch. 2.1), in memory functions (Ch. 2.2), in sensory decline (Ch. 2.3), and in multisensory processing (Ch. 2.4). These chapters help the reader to better understand the main topic of this thesis, which is the semantically congruent multisensory memory and the congruency effect. Previous research and current knowledge of this field is provided in the last chapter (Ch. 2.5) of this section.

2.1 Normal cognitive ageing

As the population continues to age at an increasing rate, it is necessary to develop a more complete understanding of the cognitive and neurological changes that accompany the ageing process. In order to conduct such research, investigators must first define normal ageing. Two views on normal ageing have been proposed: (1) the biological perspective and (2) the lifespan development perspective [33] (Fig. 1). Research conducted from the biological perspective has suggested that cognitive abilities such as memory, processing speed, and cognitive flexibility peak between 18 and 30 years of age, after which the process of normal ageing begins [34]. Thus, according to the biological perspective, normal ageing is associated with declines in cognitive domains including memory, reasoning, and spatial abilities [35].

Cognitive activity, in turn, is continuous information processing. The brain processes the information that comes from the events happening in surroundings and in the body. The vast majority of the processing is unconscious, only the part that is under the attention and in the processing of the working memory is conscious. Perception, memory and executive functions are some examples of various different cognitive functions. The key issue in perception is the separation between the relevant and irrelevant matters. Furthermore, attention and executive functions have their role in active encoding and in retrieval. [36]

Already from middle age, the ageing causes mild cognitive changes, but these age-related cognitive changes are gradual. The knowledge and the understanding of the world, the society and the relationships remain and even grow with the life experiences, as does the wisdom of life. Instead, the speed and the efficiency of the cognitive processes decline mildly. Despite of this weakening of data processing, flexible reasoning ability and the efficiency of working memory endure age-related cognitive changes. This slowdown and these working memory changes affect the effectiveness of active encoding and the speed of retrieval. Consequently, the effect on the memory processes is greater, the more the task requires spontaneous effort. [36]

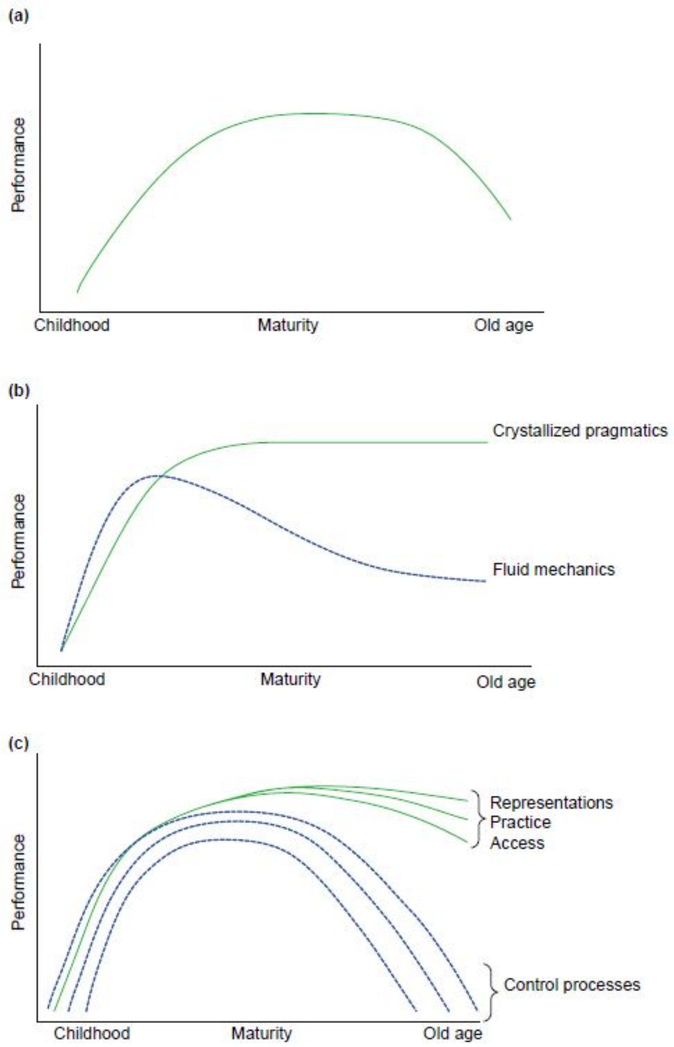


Figure 1: Three speculative models of cognitive change across the lifespan. [37]

2.1.1 Processing speed slows down with ageing

It has been shown that processing speed declines from the twenties to old age, and it has been argued that this general slowing is the primary cause of age-related declines in cognitive performance [38, 39]. Salthouse has extensively studied how processing speed develops through a lifespan and later in adulthood (see e.g. his review [40]).

Due to the decrease in brain mass resulted from declined number of synaptic contacts (see e.g. [37]), the capacity for sending nerve impulses to and from the brain decreases with ageing. As a result, the processing of information diminishes: conduction velocity decreases, voluntary motor movements slow down, and reflex times increase. Therefore, it can be reasonable to claim that older adults respond slower than young adults in different perception and memory tests.

The basic notion is that decreased speed of mental processing underlies many if not all age-related cognitive deficits, either directly (i.e. behaviour is slow and inefficient) or indirectly (i.e. by disrupting the timing of a complex sequence of mental operations). Hence, within this framework, age-related decrements in memory performance are not attributed to impaired memory processes per se, but to a generalized age difference in speed of processing. [41]

2.1.2 Age-related changes in the frontal cortex weakens cognition

The frontal cortex is essential in executive functions, as well as the role of working memory. In normal ageing, structural changes, such as neuron loss and shrinking of neurons, are greater in the frontal lobe than in other areas of the brain. This partly explains the age-related cognitive decline that is based on executive functions. As a result, the ability to maintain material temporarily in memory remains but more complex processing slows down. Thus, flexibility weakens, which can be seen broadly in different cognitive performances. [42]

Additionally, individual variations increase among elder age groups. General illnesses, lifestyles, diets and the cumulative effect of psychosocial factors in cognitive ageing have been suggested, as an explanation. [42] For example, the level of education has been found to be important mediator of the age effect on semantic and episodic memory [43].

2.2 Memory functions in normal ageing

The most commonly experienced change in cognitive functions even among healthy elders is memory impairment [41]. This is mostly due to changes in cognitive processing (as previous chapter described) which affect different memory stages. The weakening of working memory influences both the efficiency of learning and memory retrieval. As widely known, memory retention does not decline much. In different neuropsychological tests, it has been studied that free retrieval declines but the retrieval based on clues or recognition remain. [42]

Memory has different forms. On the basis of clinical and neuropsychological studies, various groups of patients may have great difficulties in certain memory tests, whereas their performances in other tests may be at the same level as that of the normal population [41]. During more recent years, brain imaging studies have confirmed such observations, by showing that certain brain structures are activated for certain tasks and other brain structures for other tasks (see e.g. a review [44] and [45]).

A rough division between different memory systems would be the division into short-term and long-term memory, with the long-term memory divided into declarative and non-declarative memory (see Fig. 2) [45]. Human memory system is mostly considered to consist of five different memory systems (proposed by Tulving [46]) but in this thesis one more memory system is added, the recognition memory. The other five separate but interacting components of memory are: working memory, episodic memory, semantic memory, procedural memory and perceptual representation system (priming). Working memory is considered to be short-term memory, and the rest are under long-term memory: episodic memory, semantic memory and recognition memory are part of the declarative memory, and procedural memory and perceptual representation system are considered to belong under the non-declarative memory. In addition to these six memory systems, there is sensory memory that captures the sensory information coming from the five different senses (sight, hearing, taste, smell and touch). The schematic drawing of human memory systems in Fig. 2 lightens the way how human memory is constructed and comprehended in this thesis.

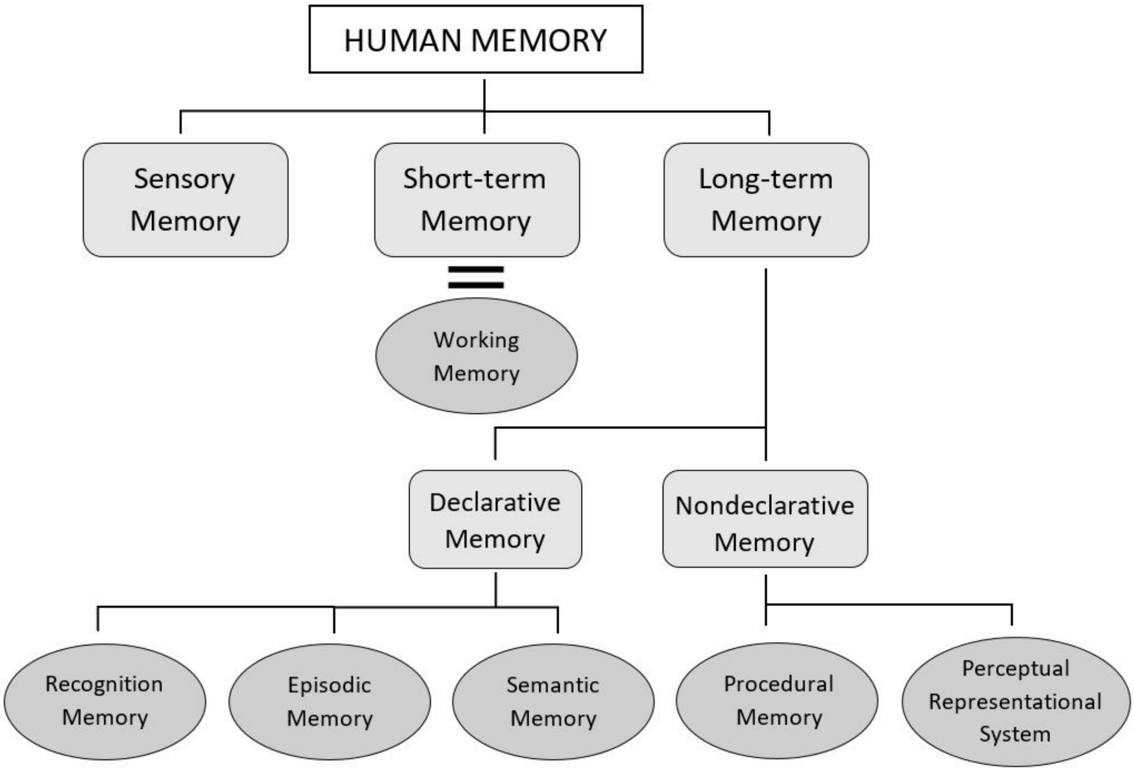


Figure 2: Schematic drawing of human memory systems as they are considered in this thesis.

2.2.1 Human memory systems

For better understanding of how these different memory systems age, it is essential to know first what these concepts mean (see e.g. [45] for more information about the topics discussed in this chapter). Starting with procedural memory, that is, the acquisition and use of various kinds of behavioural skills. Procedural memory operates at an automatic level and its output is non-cognitive. The acquisition of most procedural skills, like biking and swimming, is gradual and slow. Since this memory system's working is automatic and therefore requires almost no active effort, ageing does not affect it. Another memory system, perceptual representation (PRS), operates at an automatic and unconscious level as well. PRS is used for identifying objects in the surrounding world. Both memory systems belong under the non-declarative memory (i.e. implicit memory) (Fig. 2), where the retrieval is unconscious.

Under the declarative memory, there are semantic memory and episodic memory (Fig. 2). Semantic memory makes it possible to acquire and retain general knowledge about the world at large; and similar to procedural memory and PRS, its retrieval is implicit. It includes the meaning of words and associations between words, concepts and symbols and their associations, and facts of the world, such as Helsinki is the capital of Finland and the Earth is spherical. Episodic memory, on the other hand, is used for the encoding of personal experiences and conscious recollection of events and episodes of one's own past; it operates at a conscious level and the retrieval is explicit. This memory system, unlike the other three already mentioned, operates backwards in time, at the time of memory retrieval. In episodic memory tests the participants have to travel back in time to a given study episode in order to access the information needed. In this sense, episodic memory is strongly dependent on contextual cues for proper access to the information that needs to be remembered.

Working memory, in contrast, is a memory of the present. It is usually referred to as short-term memory (Fig. 2), especially some decades ago, but the most commonly used term today is working memory, and it will be used in this thesis. The term *working memory* is favoured because it captures the active role played by this memory system in ongoing processing information. Thus, this memory system makes it possible to hold and process information that is in the focus of consciousness. Storage is short-lived and temporary, and it operates fully at a conscious level.

2.2.2 Recognition memory

Recognition memory has been placed under declarative memory next to semantic memory and episodic memory (Fig. 2) because it has explicit storing and conscious retrieval (i.e. requiring conscious thought), which corresponds with the definitions of declarative memory (see e.g. [47]). Additionally, recognition memory shares a similar brain region with declaration memory: the medial temporal lobe [48, 49]. Classically, recognition memory has been defined as the ability to accurately assess that a stimulus has been encountered before [50]. There can be direct retrieval, which is the ability to remember a stimulus in the absence of that stimulus. There can also be discrimination components in which the learner may be able to distinguish between a stimulus that had been previously presented and a new stimulus, without any further knowledge of either one, which is the setting in the current study.

Many cognitive neuroscientists believe that the above mentioned characteristics split recognition memory into two integrated components (see e.g. [49]): *recollection* and *familiarity*. Recollection involves remembering discrete details about an experience to which the learner has been previously exposed. The second component, familiarity, is a stripped-down version of the first. This involves the conscious awareness that some object has been encountered before, but without an ability to recall anything further about it, that is, a "gut" feeling.

2.2.3 Effect of age on memory systems

The illustrative picture of how ageing influences different memory systems is presented in Fig. 3. The graph is from the paper by Nilsson [45], where he summarizes their Betula study's results [51]. Short-term memory here represents the span of the memory rather than the function of working memory, thus, there is no clear age-related deficit visible in the Betula study's results from short-term memory [51]. However, working memory, to wit, does suffer from ageing (see e.g. [52]).

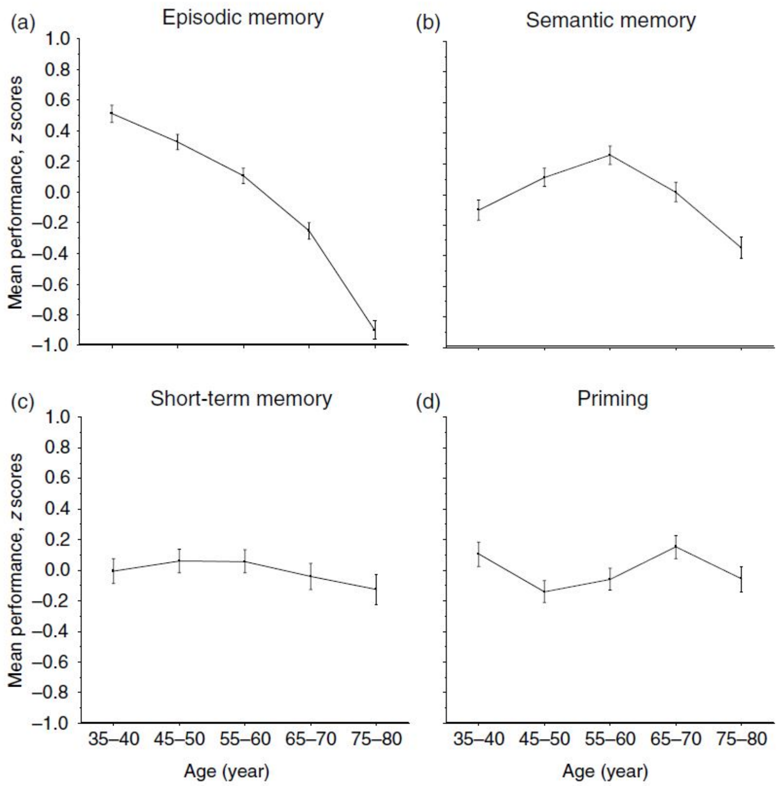


Figure 3: Illustrative picture of how different memory systems are influenced by the age. In the graphs, there is represented mean performance as a function of age in different tasks assessing episodic memory, semantic memory, short-term memory and the perceptual representation system (priming). [45]

From the declarative memory parts, there is a clear decline in episodic memory that is almost directly proportional to age. Behind the change is the functional connection with the working memory efficiency. Already back in 90s, all available evidence from cross-sectional research showed that there is a linear, decreasing memory performance as a function of age for episodic memory [53]. Longitudinal studies suggest, however, that this age deficit may be an overestimation, by showing a relatively stable performance level up to middle age, followed by a sharp decline. Semantic memory remains better than the episodic because the function of this memory type is based on the usage of previously learned memory contents. Non-declarative memory (i.e. learning skills and habits) remains because the retrieval happens automatically without conscious struggle.

Generally considering, studies on semantic memory, short-term memory, perceptual representation system, and procedural memory have shown a relatively constant performance level across the adult life span, although some tasks used to assess short-term memory and procedural memory have revealed an age deficit. It has been noticed that age-related changes resemble the ones who suffer frontal lobe damages: active memory usage causes difficulties but the automatic function remains, which also explains the differences in ageing in different memory systems.

2.3 Sensory decline in normal ageing

As it is already well known and studied, humans suffer sensory decline with ageing. In this chapter, only sensory decline in hearing and in vision are covered since these senses matter the most for better understanding of the results of the current study. The focus is only in normal ageing, and thus all the pathologies related to decline in seeing or hearing are not considered.

2.3.1 Hearing

It is very common, and widely observed among normal citizens too, that older people might not hear that well in a noisy environment and one might need to repeat the same thing again with a louder and more articulated voice in order for an older participant to hear the words. Also, many of the high-frequency sounds are missed by elder listeners, which is actually the most common age-related symptom of a hearing deficit. Certainly, this hearing loss affects the life by lowering its quality, since it affects communication. Hearing and understanding of speech in situations where there is interference gets more difficult in ageing [54, 55]. Also, cognitive skills and central auditory processing have their contribution on surviving in communication situations [54, 55].

The age-related progressive hearing loss, named *presbycusis*, has been proposed to be related to damaged and lost hair cells or degeneration of the nerve pathway for hearing [54, 55]. It is also associated with difficulties in speech discrimination, which is why hearing and understanding of speech in situations where there is interference deteriorates later in life [54]. *Presbycusis* is described as a bilateral loss of auditory sensitivity that progresses from high to low frequencies with ageing and it is caused by a primary loss of hair cells in the basal end of the cochlea [54, 55].

2.3.2 Vision

Also commonly noticed is elder people's difficulty to see close up or in the dark. Around the age of thirty, visual acuity starts to gradually become worse with age [56] and thus most of the people in their middle age wear eyeglasses. Additionally, many older adults avoid driving when it gets darker. The incidence of *presbyopia*, that is, the lenses losing their ability to curve to focus on objects that are close, and other age-related vision impairments can be expected to grow with the growing number of elderly people in the society in the following years. This is why there is a pressing need to identify the nature of age-related vision impairments, and how they impact older adults' performance of everyday visual tasks [57].

It is already widely known that several age-related changes occur in the eyes. The lens loses some of its elasticity and thus cannot change shape as easily, resulting in *presbyopia*. Therefore, older people cannot read print at the same close range as younger people. However, this normal ageing is typically corrected with glasses to enable the eye to focus for close vision. Additionally, the muscles that regulate the size of the pupil weaken with age and the pupils become smaller, react more slowly to light, and dilate more slowly in the dark. For these reasons, elderly people find that objects are not as bright, their eyes may adjust more slowly when going outdoors, and they have problems going from brightly lit to darkly lit places. Finally, as we age the sharpness of vision decreases, colour and depth perception are reduced, and "vitreal floaters" increase. (As a reference see e.g. [58] pp. 610–611.)

Another commonly observed and well established phenomenon is the slowing of visual processing, which contributes to higher-order processing problems characteristic of cognitive ageing. The slowing of visual processing among older adults is observed as a slowdown in detecting, discriminating, recognizing, or identifying visual targets. The cognitive functions participating in this can be e.g. associative learning, working memory and inhibition. Understandably, slowed processing speed in older adults has negative implications for their everyday life. Fortunately, it has been shown that visual processing speed training has led to more efficient completion of everyday visual tasks and this faster processing speed in older adults has enhanced several aspects of everyday functioning and health among this age group. [57]

2.4 The effect of age on multisensory processing

Research shows that multisensory information is processed simultaneously, such that the probability that objects and events are detected rapidly, identified correctly, and responded to appropriately is enhanced [10]. When more understanding is gained from the visual information during e.g. the speech perception, auditory and visual cues are merged into a unified percept, a mechanism known as audiovisual integration [17]. The gain the perceiver receives is called audio-visual gain.

Information from the surrounding environment through different senses is integrated by the brain to modify our behaviours and enrich our perceptions. Multisensory integration is an integral aspect of functioning and mobility in the real world. It is only through the appropriate binding and integration of information from the different senses that a meaningful and accurate perceptual shape can be generated [13]. Although a great deal is known about how such cross-modal interactions influence behaviour and perception in adults, there is also considerable evidence on how ageing impacts these multisensory processes [59].

2.4.1 Non-semantic multisensory perception

In the first studies about multisensory perception conducted with young and older adults, researchers investigated how participants perceived simple stimuli. The simple stimuli could have been single-tone beeps or lights or different colours when the stimuli were presented to more than one modality at the same time (e.g. audiovisual stimuli). The interest of these studies was to measure whether older adults perform differently compared to young adults by comparing their response times and response accuracies.

One of the first to compare multisensory enhancement between young and older adults in the audiovisual context was Laurienti et al. [13]. They examined these age groups' speed of discrimination responses to the presentation of visual, auditory or combined audiovisual stimuli. In the task, the participants were asked to differentiate between different colours by pressing a certain button. The results revealed that audiovisual stimuli speeded response times in both age groups but the performance gain was significantly greater in the older adults than in the younger ones. The results revealed also that older adults exhibited a greater peak and a broader temporal window of multisensory enhancement than young adults. As explanations for their findings, Laurienti et al. suggested generalized slowing with ageing, speed-accuracy trade-off, and a broader window over which multisensory benefits could occur. They also discussed that a decline in each of the unisensory modalities could have played a role since the major benefit of multisensory integration was seen when unisensory performance levels were low.

Since enhanced integration may be explained by alterations associated with general cognitive slowing, as Laurienti et al. [13] also suggested, Peiffer et al. [14] utilized a task that eliminated most high-order cognitive processing. They used a simple audiovisual detection task where the participants, young and older adults, were instructed to react as quickly as possible when they either saw or heard any stimulus. The visual stimuli were green light emitting diodes (LED) and the auditory stimuli white noise. There were visual alone, auditory alone, and multisensory trials (auditory and visual stimuli occurring simultaneously). As a result, no significant differences in unisensory response times were seen, however, older adults experienced greater multisensory gains than younger adults. In the multisensory trials, older adults actually tended to respond faster than younger adults. These results from Peiffer et al. suggested that the increased integration evident in older adults is not an epiphenomenon of general cognitive slowing, but rather is due to basic changes in the multisensory processing stream. Therefore, data from Peiffer's et al. study support unique age-related changes in older adults for the sensory processing of multisensory information.

Mahoney et al. [15] studied multisensory integration effects with a simple reaction time task in young and older adults by comparing reaction times across paired sensory inputs, e.g. audiovisual stimuli. The results revealed that participants, old and young adults, were significantly faster at responding to multisensory compared to unisensory stimuli which indicated successful neuronal integration of information across sensory systems in both old and young adults. Mahoney et al. found as well that age did not interact with sensory modality type within the multisensory or the unisensory conditions, suggesting similar sensory processing across both age groups. This finding contrasts with findings from Peiffer et al. [14] and Laurienti et al. [13], where older adults demonstrated greater audiovisual multisensory enhancements than younger adults. Additionally, reaction times to all sensory conditions were significantly slower for older compared to young adults. Mahoney et al. suggested that in their study differences in multisensory relative to unisensory processing across age groups were likely indicative of differences in sensory processing at the synaptic level between older and young adults. On the other hand, Mahoney et al. got the same result as Laurienti et al. that multisensory stimuli speeded response times in both age groups [13].

These results from Laurienti et al. [13], Peiffer et al. [14] and Mahoney et al. [15] all were in line with the fact that both young and older adults benefited from the multisensory conditions. However, Mahoney et al. did not find the multisensory gain for older adults that Laurienti et al. and Peiffer et al. reported. Another discrepancy was that Peiffer et al. found faster response times for older adults, but in the study from Mahoney et al. older adults had slower response times than younger adults. As a conclusion, it can be assumed that older adults will benefit from multisensory conditions, and possibly then show faster reaction time responses.

The aim of these studies was to examine how older adults integrate multisensory information and therefore, they used simple stimuli like lights or circles, and tones or white noise that barely require any cognitive processing. However, this is not that well comparable with the real world where many different inputs through the senses are received. Thus, there are many studies that have investigated how older adults perform when they are asked to recognize speech. The next chapter covers current literature about audiovisual speech perception that includes semantic context.

2.4.2 Audiovisual speech perception

Speech perception is a particularly studied domain in older adults due to its importance for communication and the implications of speech comprehension for social interactions. In the audiovisual speech perception study designs, in most cases there is a talker, a male or a female, who has been recorded from shoulders up while she or he is articulating syllables, words, or sentences that participants need to recognize or repeat. In the examination, the studies have applied McGurk effect [16, 17, 19], congruency effect [17–19], and response accuracy [18, 19, 21, 60–63].

McGurk effect occurs when incongruent audiovisual stimuli are integrated successfully

McGurk & MacDonald [64] understood that most of the verbal communication occurs in a situation where the listener can hear and see the speaker. They had observed that when a video of a person's talking head, in which repeated expressions of the syllable /ba/ had been dubbed on to lip movements for /ga/, was presented to normal adults, they reported hearing the syllable /da/. However, when the syllables were presented alone (only visual or only auditory), the adults reported hearing the syllables as they were repeated, accurately. Nevertheless, the main finding of McGurk & MacDonald's study was that a role for vision (perceiving lip movements) in the perception of speech by normally hearing people is clearly illustrated, that is, the McGurk effect. Taken together, a "fused" response occurs when the information from different sensory modalities is integrated successfully into something new with an element not presented in either modality but affected by both of them.

Thus, the existence of the McGurk effect provides further evidence that auditory and visual information interact in the perception of speech. Cienkowski & Carney [16] discussed that the use of conflicting auditory and visual speech stimuli might provide a test of integration performance for older adults. The results from Cienkowski & Carney [16] showed that at the syllable level on average, older adults are as successful as young adults at integrating auditory and visual information for speech perception. However, the difference between these age groups were found in the chosen response alternatives since older adults lean toward visual alternatives, whereas their younger counterparts toward an auditory alternative. The number of fused responses, however, did not differ across the ages tested.

Huyse, Leybaert & Berthommier [17] gained similar results as Cienkowski & Carney [16] when they tested audiovisual integration with visual clear condition; the audiovisual gain was similar in both age groups, that is, the older adults did not differ from the younger adults in their audiovisual integration abilities. However, the difference occurred when the visual cue was reduced and the noise was stationary. In this case the audiovisual gain was weaker in older adults. If the auditory cue was clear, older adults managed to compensate for the loss of lipreading abilities by using the auditory information available. Huyse et al. [17] also found out that the significant impact of ageing on audiovisual integration is that older adults put more weight to the auditory input than younger adults when the visual input is degraded. Taken together, ageing had a significant impact on audiovisual speech integration when the visual speech cue was less informative, but not when it was clear. Thus, according to Cienkowski & Carney [16], older adults trust more visual cues, and according to Huyse et al. [17], older adults trust more auditory cues when the visual cues are degraded.

Sekiyama et al. [18], from their behalf, wanted to examine whether older adults with normal hearing use visual information more than young adults. The results followed the findings from the study of Cienkowski & Carney [16], since it was found that the older adults were more strongly influenced by visual speech than the younger ones when the SNR of auditory speech was acoustically identical. Additionally, Sekiyama et al. [18] found that there were no age-related differences in visual-only lipreading accuracy. They concluded that the enhanced visual influence for the older adults is likely to be associated with an age-related delay in auditory processing.

Setti et al. [19], on the other hand, found that audiovisual integration of incongruent audiovisual words was higher in older adults than in younger adults. Older participants also recalled more illusory audiovisual words in sentences than younger adults. Setti's et al. findings suggested that the relatively high susceptibility to the audiovisual speech illusion in older participants was due to more perceptual than cognitive processing. Similar conclusions were made by Sommers, Tye-Murray & Spehar [60] as well, since their findings suggested that the poorer performance of older adults in the audiovisual condition was a result of reduced lipreading abilities (perception) rather than a consequence of impaired integration capacities (cognition).

Observing how older adults are affected by the signal content

Many audiovisual speech perception studies have tested older adults ability to integrate audiovisual stimuli in varying conditions and with different signal content. Tye-Murray et al. [61] compared the ability of young adults and older adults to integrate auditory and visual sentences under conditions of good and poor signal clarity. In their investigation, they applied the Principal of Inverse Effectiveness (PoIE) that predicts that both young and older adults will show enhanced integration of auditory and visual speech stimuli when these stimuli are degraded, and especially older adults would show enhanced integration during audiovisual speech recognition relative to young adults. However, the results showed the opposite; neither the auditory enhancement nor integration enhancement measures increased when signal clarity in the auditory or visual channel of audiovisual speech stimuli was decreased, nor was either measure higher for older adults than for young adults. These results also suggested that ageing did not affect integration enhancement when the visual speech signal had good clarity, but may have affected it when the visual speech signal had poor clarity.

Tye-Murray et al. [20] have also studied cross-modal enhancement of speech detection in young and older adults. The results showed a different pattern of cross-modal enhancement for older adults than for young adults. Whereas the young adults benefited from every audiovisual condition older adults benefited only from the high-contrast video of the talker's face, the performance in other two audiovisual conditions were similar to the baseline condition (auditory-only). They suggested that signal content affect cross-modal enhancement and proposed that ageing undermine the ability to take advantage of degraded visual speech information. Tye-Murray et al. concluded that "older adults with apparently normal vision may still have a deficit or deficits under suboptimal viewing conditions, which preclude their benefiting from visual signals that produce cross-modal enhancement in young adults".

Maguinness et al. [21] investigated the efficiency of auditory and visual integration in an older population by manipulating the relative reliability of the auditory and visual information in speech. The results showed that older adults had better recall for sentences when audiovisual condition had no blur than when it had, especially the non-meaningful sentences were better recalled in audiovisual no blur condition compared to audiovisual blur condition. Generally younger adults performed better than older adults, and especially non-meaningful sentences were recalled better by younger ones than older adults. The multisensory enhancement were greater for older adults when the sentences were non-meaningful than when the sentences were meaningful. As Maguinness et al. explained, their findings suggested that information with unfamiliar or complex content will be better remembered if presented in an audiovisual format, where information from both sensory components is reliable.

Alm & Behne have investigated, whether age-related audiovisual experiences affect audiovisual asynchrony perception [65], and the relationship between cognitive processing speed and audiovisual asynchrony detection in speech [66] by comparing young and middle-aged adults. The middle-aged adults were expected to have a narrower synchrony window than young adults, and as expected, middle-aged adults showed less tolerance for audio-lead than young adults [65]. Additionally, they found that cognitive processing speed influenced audiovisual asynchrony detection in speech in young adulthood, and that this influence was considerably reduced by middle adulthood [66].

Some neuroimaging data about audiovisual speech has been reported as well. Winnike & Phillips [62] recorded Event-Related Potentials (ERP) from the brain while participants were carrying out a behavioural experiment. In behavioural tests both age groups revealed an equivalent behavioural audiovisual speech benefit over unisensory trials. On the other hand, the ERP data revealed that older adults showed more clear facilitations of neural responses on audiovisual speech trials than their younger counterparts. These results showed that audiovisual speech processing was not only intact in older adults, but that the facilitation of neural responses occurred earlier in and to a greater extent than in younger adults. This is probably the reason why it seems that older adults benefit more from additional visual speech cues than younger adults since they might try to compensate changes in the sensory ageing.

Summarizing the outcome of audiovisual speech perception studies

Taken together, it has been shown that older adults put more weight on the visual cues [16–18]. However, when the visual cue is degraded, older adults compensate the loss by relying more on the clear auditory cue [17]. Hence, the enhanced visual influence was concluded to be associated with an age-related delay in auditory processing [18]. On the other hand, older adults have high susceptibility to audiovisual speech illusion since they repeate more illusory audiovisual words [19, 60]. This was suggested to be more due to perceptual than cognitive processing. Also, age does not influence integration enhancement when visual speech signal have poor signal clarity and the enhancement does not increase with a decreased signal clarity [61]. Additionally, older adults benefit only from high-contrast video, and thus, it was suggested that signal content affects audiovisual enhancement, and ageing undermines the ability to take advantage of degraded visual speech information [20].

Older adults gain greater multisensory enhancement when they need to indicate non-meaningful sentences and therefore, it was suggested that information with unfamiliar or complex content would be better remembered when presented audiovisually [21]. Also in neural responses, more clear facilitation has been found for older adults, since neural responses occur earlier and to a greater extent [62]. With middle-aged adults less tolerance to audio-lead has been detected compared to younger adults [65], and that the influence of cognitive processing speed on audiovisual asynchrony detection in speech is reduced by middle adulthood [66]. Generally, older adults suffer more from degraded signal clarity and they integrate more easily confusing audiovisual content than their younger counterparts, that is, older adults seem to show greater McGurk effect than younger adults because they are more prone to lean on the visual cues over the auditory cues.

2.4.3 Supplementing speech understanding with visual text

Older adults have been generally observed to have challanges understanding in difficult listening conditions possibly due to their degraded hearing and vision. Therefore, Krull & Humes [67] tested the hypothesis that partially accurate visual text from an automatic speech recognizer could be used successfully to supplement speech understanding in difficult listening conditions in older adults, with normal or impaired hearing. Krull & Humes did perceptual tests and cognitive tests to their participants to assess verbal comprehension, processing speed, and working memory.

The results showed that young and older groups performed similarly for all perceptual measures. Significant differences were found in cognitive measures. Significant age effects were noted e.g. for working memory, processing speed, perceptual speed, and cognitive speed. Krull & Humes found that although both young and older adults benefited from partially accurate text combined with degraded speech, the benefit changed as a function of text accuracy and background noise. In both younger and older adults, cognition emerged as a key predictor of speech-text integration ability because two cognitive factors best predicted performance for this ability.

2.5 Semantically congruent multisensory memory and congruency effect

A quite wide variety of studies in the field of auditory and visual recognition memory, where effects of semantically congruent multisensory experiences have been tested, exists. The stimulus is said to be semantically congruent with another stimulus if they share same semantic meaning and can be paired to origin from the same object or material (e.g. hearing a dog's barking and seeing a dog). Stimuli are incongruent among themselves already if they do not origin from the same object although they would share the same semantic category (e.g. hearing a cat's meow but seeing a dog). Since information is integrated spatially, temporally and semantically coherently, and such congruent multisensory information facilitates our ability to perceive the world around us [10], it has been studied whether the multisensory information can improve the ability to remember things too (see e.g. [27]). The most common outcome of these studies has been the congruency effect, that is, the memory performance has been enhanced by a semantically congruent multisensory stimulus [22–30], except [31]. In most of the studies audiovisual stimulus has been used [24, 25, 27–30], but some have used somatosensory-visual as well (see e.g. [23]).

Many of the studies have applied continuous recognition memory test, where the participants are responding during the experiment, discriminating between items seen for the first or for the second time during a block of trials [22–25]. In these tests, most commonly visual stimuli have been used as a task-relevant, that is, a "to-be-remembered stimulus", and auditory stimuli as a task-irrelevant, that is a possibly supportive material. Nevertheless, there are also some studies with auditory stimulus as a task-relevant stimulus [24, 25]. The first appearance of the stimulus could have been unisensory or multisensory but the second appearance is every time only unisensory representing the task-relevant stimulus. The participants need to keep the object in their mind for a relatively short period.

Another approach for assessing recognition memory performance are the two-part ("old/new") recognition memory tests with separated experimental phases: study phase (encoding) and test phase (retrieval). The encoding part can be unisensory or multisensory but the retrieval is always unisensory [26–32]. So far, the most of the multisensory memory studies have been conducted with young adults (e.g. [22, 27, 28, 68]), and one with school-aged children [30]. According to the current knowledge, there are little multisensory memory studies conducted with older adults.

A decade ago, Luo, Hendriks & Craik [32] studied different memory encoding conditions that could improve recollection for both young adults and older adults. In the study phase, participants had to watch visually presented lists of words or listen to those lists of words. Participants were instructed to memorize both the words and the modalities (auditory or visual), in order to perform well in the test phase. For testing the memory encoding in recognition memory, they manipulated the encoding condition, however, only in the visual lists. The manipulation was expected to improve the performance later in the test phase. These different manipulations were words alone (baseline), with pictures (dual-visual), as word fragments (generation condition), or with sound effects (audiovisual). Since the purpose of this study was to find the way to improve the performance, the encoding manipulation was always semantically congruent. As a result, Luo et al. observed that older adults can only benefit from pictures (dual-visual) and word fragments presented with the visual lists, and that sound effects (audiovisual) enhance recollection only in young adults but not in older adults. Luo et al. argued that older adults are less effective with integrative processing that sound effects would have demanded, and thus the improvement of recollection in younger adults is to a greater extent. However, according to these findings, it is not known yet whether older adults could benefit from other audiovisual stimulus combinations.

2.5.1 Continuous recognition tasks in evaluating semantic congruency effect

Murray and co-workers [22, 23, 25, 68] and Moran et al. [24] have particularly used continuous recognition tasks to evaluate the semantic congruency effect on multisensory events. In the studies, the items were line-drawings or sounds [22, 23, 25], or coloured sketched drawings and sounds [24]. Before, in the studies of Murray and co-workers [22, 68], the task-relevant item was solely visual stimulus (i.e. line drawings), however, in recent years one paper has been published from Murray and co-workers [25] and one from Moran et al. [24] with the auditory stimulus as the task-relevant item. In these continuous recognition tasks semantically congruent, incongruent and non-semantic pairings have been randomly varied during the block of trials.

These studies reported significant main effect for semantically congruent pairings when the task-relevant stimulus was appearing for the second time (old stimulus) in the block [22–25]. This reference to the performance was more accurate when the task-relevant stimulus had been presented with semantically congruent task-irrelevant stimulus in the first appearance during the block. This congruency effect happened for visual task-relevant stimulus as well as for auditory task-relevant stimulus. In addition, both Moran et al. [24] and Thelen et al. [25] found that recognition of auditory task-relevant stimulus benefited more from visual task-irrelevant stimulus than the reverse. Moran et al. [24] additionally stated that difficult sounds (auditory) were easier to recognize when they were combined with a supportive picture (visual), and that the recognition of easy pictures did not benefit from sounds as much as from pictures. Murray et al. [22] and Lehmann & Murray [23] found also that mean reaction times for visual-only trials were faster than for semantically congruent audiovisual stimuli. However, Moran et al. [24] reported that mean reaction times for auditory-only were slower than for multisensory recognition conditions.

2.5.2 Two-phased recognition memory tests in evaluating semantic congruency effect

Recently, Heikkilä, Alho, Hyvönen & Tiippana [27] showed that semantically congruent multisensory information enhances memory performance. They had five different study blocks with different stimuli in every block. Depending on the block, the stimuli were audiovisual or dual-visual (two visual stimuli together) and combinations of spoken or written words, natural sounds, or pictures. In some of the blocks participants were asked to ignore auditory stimulus and in the rest the visual stimulus. In the audiovisual blocks where the auditory stimuli were to-be-remembered, natural sounds were accompanied with pictures or spoken words were accompanied with written words. The participants recalled the sounds or the spoken words while ignoring the pictures or the written words. The blocks also included different semantic congruency conditions: semantically congruent, semantically incongruent, and as a control condition there was a non-semantic condition that was white noise (auditory), a frame of white noise (visual) or a row of six letters of X (visual). These different conditions alternated randomly in the study block. This study revealed a significant main effect for semantically congruent audiovisual stimulus compared to the non-semantic condition in both blocks (sounds with pictures and spoken words with written words). Thus, participants performed better in the auditory recognition memory test when they had semantically congruent visual stimuli in the encoding.

Heikkilä & Tiippana have also studied semantic congruency among school-aged children [30]. The study design resembled the one from their previous study [27], except now they had one experimental setting more: spoken words were accompanied with pictures. They also had three different congruency conditions: semantically congruent, semantically incongruent, and non-semantic as a control condition. As expected, the experiment of sounds with pictures showed similar results to their previous study [27], that is, semantically congruent audiovisual stimulus enhanced memory performance when the children were asked to recall the sounds in the test phase. However, the children did not benefit from written words when recalling spoken words, the resuts were inconsistent when compared to their previous results [27] since young adults had been benefiting from written words when recalling spoken words. On the other hand, school-aged children benefited from pictures when recalling spoken words, that is, pictures enhanced auditory recognition memory performance in children. This was not tested with young adults [27].

Since there is a considerable amount of evidence that congruent multisensory experiences facilitate subsequent unisensory recognition memory performance (see e.g. [27, 68]), Heikkilä et al. [29] wanted to investigate whether simultaneous presentation of two semantically congruent stimuli during encoding enhance later retrieval also for congruent dual-visual stimuli. Their aim was to study the effect of task-irrelevant stimuli on the memory performance of the task-relevant stimuli. The results demonstrated that both multisensory and unisensory semantic congruency can facilitate recognition memory. Thus, the congruency effect is not solely multisensory.

Very recently, a study was published about the benefits of audiovisual semantic congruency in the recognition memory performance for sounds, spoken words and written words. The main interest of Heikkilä, Alho & Tiippana's study [28] was to study which kind of semantically congruent audiovisual stimulus pairs can facilitate the precision of recognition memory in relation to unisensory stimuli. The study included three blocks: in one block participants recalled sounds, in another block spoken words, and in the last one written words. These task-relevant stimuli were then accompanied with pictures, with written words or with spoken words, or presented alone (unisensory). The results revealed statistically significant difference for sounds with pictures and for sounds with written words compared to sounds-only. Similarly spoken words with pictures were remembered significantly better than spoken words alone. However, spoken words with written words did not show any significant difference compared to unisensory situation. Additionally, written words presented with spoken words or alone did not show any significant difference. Heikkilä et al. [28] observed that recognition memory performance for sounds and spoken words enhanced with semantically congruent visual information, except when both auditory and visual items were verbal.

There are also conflicting results from semantically congruent recognition memory. Cohen, Horowitz & Wolfe [31] did not find significant main effects for audiovisual conditions compared to sounds or verbal description alone. They only found a significant main effect for pictures-only comparing to the rest of the conditions; the recognition memory performance for pictures was significantly better than the recognition memory performance for any other condition. In their four experiments with all together 11 different conditions, they had different participants: 12 participants for each condition, while in Heikkilä's et al. studies [27, 28, 30] the same participants completed all the different settings with different conditions. In Cohen's et al. study participants recalled the auditory stimulus in every condition. However, only two of the conditions were audiovisual: sounds with pictures, and sounds with verbal description. Other conditions were solely sounds, pictures, verbal descriptions, music, and spoken language. One possible explanation for this contradictory result that is also stated by Heikkilä et al. [27], is the presentation time of the sounds which was full five seconds comparing to those 400 milliseconds that was the presentation time of sounds and 350-780 milliseconds of spoken words in Heikkilä's et al. studies [27, 28, 30]. In the Cohen's et al. study, the participants had more time to process the sounds, and thus the audiovisual semantic congruency did not help participants to remember the sounds any better than in the unisensory auditory condition.

A study has also been published that used magnetoencephalography (MEG) to examine the latency of modality-specific reactivation in the visual and auditory cortices during a recognition task. Ueno et al. [26] wanted to determine the effects of reactivation on episodic memory retrieval since it has been reported that reactivation of sensory information pathways is associated with the process of long-term memory retrieval. During the encoding, participants (young adults) were asked to memorize the encoding words that were accompanied either with a sound (audiovisual), with a picture (dual-visual), or no additional item (unisensory). The results from behavioural data indicated that the recognition of written words was higher in cases of auditory (audiovisual) than visual (dual-visual) encoding, and also for visual (dual-visual) encoding compared with unisensory encoding. Additionally, MEG amplitudes were greater during the audiovisual condition and dual-visual condition than the unisensory condition. These results suggested that reactivation of visual and auditory brain regions during recognition binds language with modality-specific information and that reactivation enhances the performance in recognition.

2.5.3 Current study

As all here reviewed studies showed, participants benefited the most from semantically congruent audiovisual conditions compared to unisensory conditions. Additionally, multisensory conditions were generally observed to allow better memory performance compared to unisensory conditions among young participants [22, 24–28], and among children [30]. Only Cohen et al. [31] could not find this effect. In addition, as previously described, older adults did not benefit from accompanied sound effects (audiovisual condition) unlike young adults [32]. Many studies also observed that visual cues helped to remember auditory stimuli better than auditory cues helped visual stimuli [24, 25, 27, 28, 30], which could be explained by the finding that auditory recognition memory is inferior to visual recognition memory [31]. Additionally, in some of the continuous recognition tasks [22, 23] mean reaction times seemed to be slower for multisensory trials than for unisensory trials, but contradictory results were obtained too [24]. In two-part recognition memory tests mean reaction times have been reported only by Heikkilä et al. [27] in a table with a note that there were no statistically significant differences in reaction times between congruency conditions, and by Ueno et al. [26] in a table. Others (see e.g. [28, 30] have not even reported mean reaction times which might have revealed that there were no interesting findings from that point of view either.

Thus, according to the previous literature and what have been already learned, the current study examined the recognition memory performance in older adults with a two-part recognition memory test with the exact same study design as Heikkilä et al. [28]. The task-relevant stimuli were auditory stimuli, and the task-irrelevant stimuli were visual stimuli. In the study phase, i.e. during the encoding, audiovisual or unisensory stimuli were showed to the participants but in the test phase, i.e. during the memory retrieval, only auditory stimuli were presented. Semantic non-verbal and semantic verbal material were applied as they closely corresponded to real life situations, and due to the strong evidence that participants benefit the most from semantic congruency. Older adults were chosen because now, due to rapidly ageing population, it is important to find ways how multisensory encoding could enhance later memory performance.

It was hypothesized that older adults will benefit from multisensory memory cues in a similar way to young adults it has been shown in the previous literature. Additionally, it was hypothesized that no pronounced effect will be seen in mean reaction times since in previous recognition memory studies conducted with young adults, since mean reaction times have not been even reported in the articles. The results from the continuous recognition tasks, on the other hand, have been dissonant. Moreover, the study setting in the current study is different compared to the continuous recognition task studies. It was also hypothesized that sounds will be remembered more poorly than spoken words, as did young adults in Heikkilä's et al. very recent study [28]. The following questions were asked: How much do older adults benefit from pictures in the encoding? Can older adults benefit from written words during the encoding of spoken words?

3 Research material and methods

In the present study, the same experimental study design was utilized as Heikkilä et al. [28], except the participants recalled only auditory stimuli during the whole experiment. In Heikkilä's et al. study [28] there was one additional block where written words were presented alone or together with spoken words. This study block was left out from the current study since young adults did not show any significant difference between the two conditions when recalling written words. Thus, it was hypothesized that similar results could be gained with older adults which would not add anything to the study.

3.1 Participants

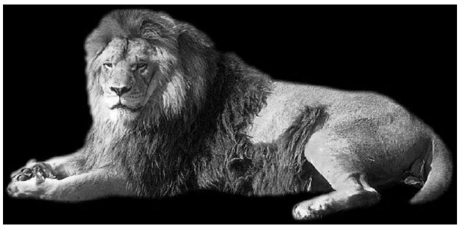
Fifty-five volunteers participated in the experiment, however, 13 of them were excluded due to seeing or hearing deficits. The remaining 42 (15 men) aged 65-85 (mean 71 years, SD 4.4 years) reported normal or corrected to normal vision and normal hearing and no dyslexia or neurological illnesses. Four participants reported Swedish as their mother tongue but having excellent Finnish. Thus the majority had Finnish as their mother language. All participants signed a written consent form (see Appendix C). The research had ethical approval from the University of Helsinki Review Board in the Humanities and Social and Behavioural Sciences.

3.2 Stimuli

Pictures of natural objects, sounds of natural objects, written words and spoken words were used as stimulus material. The inter-stimulus interval was 800 ms and the volume intensity was 50 dB.

The pictures were 76 photographs obtained from the Multimodal Stimulus Set [69] or from the internet and modified to resemble those in the Multimodal Stimulus Set (see Stimulus List in Appendix B). The photographs presented objects from several semantic categories (animals, tools, vehicles, musical instruments, and household items). They were converted into gray scale images and presented on a black background. They were positioned centrally on a computer monitor located 60 cm from the participant. Their sizes varied between 1 cm and 20 cm horizontally and vertically. In Fig. 4, two example pictures are presented from the stimulus set and how they were presented in the examination to participants.

The sounds were 150 recordings of complex sounds of objects from several semantic categories (animals, tools, vehicles, musical instruments, and household items). They were obtained from the Multimodal Stimulus Set [69] or from the internet (www.findsounds.com) and edited to match those from the Multimodal Stimulus Set. The sound duration was 400 ms.



(a) Lion with the lion's roar

(b) Flag with the word "flag" spoken

Figure 4: Example pictures with the auditory stimulus described from the stimulus set presented in the study.

The written words were 76 common two-syllable Finnish nouns from various semantic categories (animals, tools, foods, plants and household items etc.). They were presented in white Times New Roman, 40-point font size in the centre of the computer monitor. The spoken words were 306 common two-syllable Finnish nouns from different semantic categories (animals, tools, foods, plants and household items etc.). They were spoken by a female speaker, recorded, and edited to a constant intensity of 50 dB. Their duration varied between 350 and 780 ms.

3.3 Design and procedure

The study included two blocks and in both blocks there were three similar conditions: auditory stimulus alone, auditory stimulus with a picture, and auditory stimulus with a written word. These formed altogether six different tests that were presented to all the participants in random order. All tests consisted of two parts: (1) an encoding phase consisting of audiovisual or auditory (unisensory) items, followed by (2) a recognition memory task consisting of only auditory (unisensory) items. The experimenter sat next to the participant during the experiment to ensure proper concentration on the task and that the participant maintained their gaze on the screen. The study design is visually presented in Fig. 5.

During encoding, the participants were instructed to memorize the stimulus of the auditory modality while ignoring the stimulus in the visual modality. For each four audiovisual tests, there were only congruent items. The congruent items were semantically matching: for example, a picture of a dog and the sound of barking, or the same noun (e.g. "aunt") written and spoken. There were no incongruences at any point. Two of the tests had only auditory items without any visual support.

In the block 1, participants memorized 25 sounds, and in the block 2, they memorized 51 spoken words. All items were presented in random order with simultaneous onsets of their auditory and visual components. During the inter-stimulus interval, a black screen with a white fixation cross in the centre was presented.

The recognition memory task (retrieval) immediately followed the encoding phase. In this retrieval phase, all the auditory stimuli from the encoding phase were presented again with an equal number of new stimuli added (25 with sounds and 51 with spoken words). The presentation order of these stimuli was randomized and no visual support was given. The participant's task was to decide whether or not each stimulus had been presented in the encoding phase by pressing one of two keys that were the right and the left mouse buttons on a laptop keyboard representing answers *yes* or *no*. With every other participant, the *yes* was the right button and with the other half the *yes* was the left button to avoid any bias resulting from this. The next item was presented 1000 ms after the response.

All stimuli were different across the tests: for example, the concept "dog" was presented only once (either as a spoken or a written word, or as a sound or a picture) during the course of the experiment. The number of items from different semantic categories was quite evenly distributed across congruency conditions. See Appendix B (containing complete lists of all stimuli) for details.

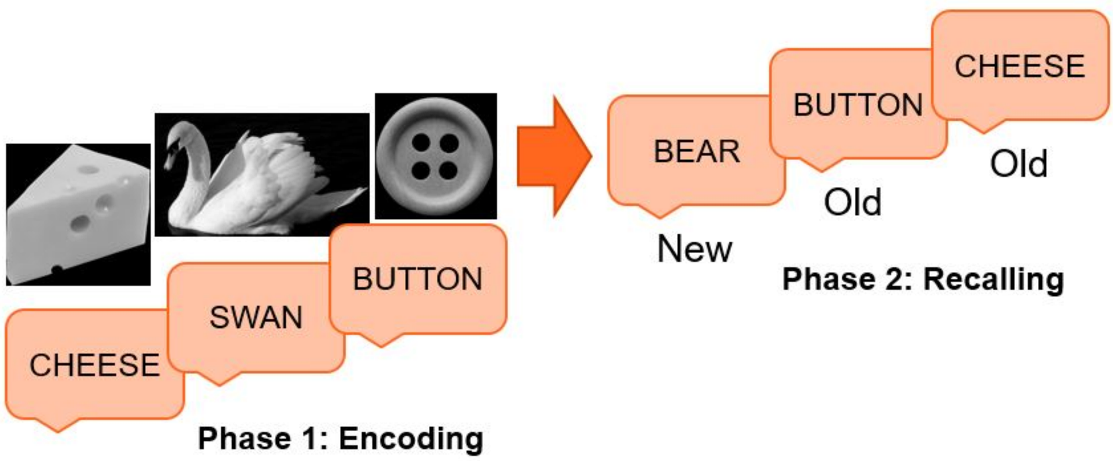


Figure 5: The study design described by a test where participants encode spoken words with semantically congruent pictures. The schema shows the two phases of the recognition memory test and how the retrieval is always auditory alone.

Block 1: Sounds

Test 1: Sounds-only

In the encoding phase, sounds were presented without any other stimuli in another modality. The task was to memorize the sounds. In the auditory recognition memory phase, the participant decided whether or not each sound had been presented in the encoding phase.

Test 2: Sounds with pictures

In the encoding phase, sounds were paired with semantically congruent pictures. The task was to memorize the sounds, while ignoring the pictures. In the auditory recognition memory phase, the participant decided whether or not each sound had been presented in the encoding phase. No pictures were presented in this retrieval phase, only a white fixation cross was present on the screen.

Test 3: Sounds with written words (text)

In the encoding phase, sounds were paired with semantically congruent written words (text). The task was to memorize the sounds, while ignoring the text. In the auditory recognition memory phase, the participant decided whether or not each sound had been presented in the encoding phase. No text was presented in this retrieval phase, only a white fixation cross was present on the screen.

Block 2: Spoken words

Test 4: Spoken words-only

In the encoding phase, spoken words were presented without any other stimuli in another modality. The task was to memorize the spoken words. In the auditory recognition memory phase, the participant decided whether or not each spoken word had been presented in the encoding phase.

Test 5: Spoken words with pictures

In the encoding phase, spoken words were paired with semantically congruent pictures. The task was to memorize the spoken words, while ignoring the pictures. In the auditory recognition memory phase, the participant decided whether or not each spoken word had been presented in the encoding phase. No pictures were presented in this retrieval phase, only a white fixation cross was present on the screen.

Test 6: Spoken words with written words (text)

In the encoding phase, spoken words were paired with semantically congruent written words (text). The task was to memorize the spoken words, while ignoring the text. In the auditory recognition memory phase, the participant decided whether or not each spoken word had been presented in the encoding phase. No text was presented in this retrieval phase, only a white fixation cross was present on the screen.

The experiment was conducted in different locations depending on where the participant preferred to perform the experiment. This was done to lower the threshold to participate in the study. The options were their own homes or university's spaces. The rooms were chosen to be quiet, dimly-lit and closed rooms with as little distraction as possible. The participant sat 60 cm from the laptop screen, where the visual stimuli appeared. The laptop used in the experiment was HP Elitebook 8460p LH1-IBS-088 with Windows 7.

The auditory stimuli were presented at approximately 50 dB(A) via headphones (Sennheiser HD201). The stimuli were presented and the response data gathered with Presentation software (Neurobehavioral systems). The right and left mouse buttons of the laptop's keyboard were used for data acquisition. Before the experiment, the participants read written instructions and performed a short practice session. The six tests (lasting about 5-10 minutes each) were presented to each participant in random order.

3.4 Data analysis

The data analysis of this study was conducted with repeated measures ANOVA. In order to gain trustworthy results, signal detection theory was applied. In this section, first, the analysis of mean reaction times is presented (Ch. 3.4.1), then the signal detection theory is provided (Ch. 3.4.2), and last, repeated measures ANOVA with its assumptions are described and its use is justified (Ch. 3.4.3). In the final chapter the data analysis procedure is described (Ch. 3.4.4).

3.4.1 Mean reaction times

Reaction times were measured by the Presentation Software applied in the study design while participants were responding *yes* or *no* in the test phase (during the retrieval). The single reaction time represented the time the participant used to answer to one signal trial. The mean reaction time of one test was obtained by taking the average of all the reaction times (over every participant) gotten in that test. In the current study, the mean reaction times for the hit rates, that are, the probabilities of responding *yes* on signal trials (stimuli presented in the encoding phase), was under the main interest.

3.4.2 Signal detection theory

In the experiment, the participants made the choice between the answers *yes* and *no* according to whether or not the stimulus was presented before. Consequently, every participant had his own criterion for answering and it could have been high or low depending on the certainty of knowing the correct answer, the level of the difficulty in the ongoing task, or the motivation, resulting in uncertainty in decision making. Due to this uncertainty in decision making, signal detection theory (SDT) was applied. In SDT this *criterion* is marked with Roman letter *c*. Another SDT measure was utilized as well, *decision making variable d'*, since there was a chance of a response bias in the current study. *d'* is stated to be unaffected by response bias since it removes the effect of fluctuation of *c*, thus making the results more comparable and trustworthy.

In the current study the tests were a type of *yes/no* tasks. Stanislaw and Todoro [70] described the application of signal detection theory (SDT) in *yes/no* tasks. In this paper, *the criterion* is defined as a value of *a sufficiently high decision variable.* On this value the participants based their response during each trial and if the decision variable was sufficiently high during a given trial, the participant responded *yes*, but otherwise *no*. Therefore, more precise definition of c would be the distance between *the criterion* and *the neutral point*, where neither response would be favoured: participants responding *yes* when the decision variable exceeds the criterion and *no* otherwise. If the criterion is located in the neutral point, c gets a value of 0. Deviations from the neutral point are measured in standard deviation units. Negative values of c signify a bias toward responding *yes*, whereas positive values signify a bias toward the *no* response.

In the current stydy, a *yes* response is made only if the stimulus item seems sufficiently familiar. On signal trials (stimuli presented in the encoding phase), *yes* responses are correct and are termed *hits*. On noise trials (stimuli presented only in the recalling phase), *yes* responses are incorrect and are termed *false alarms*. *The hit rate* (HR), the probability of responding yes on signal trials, and *the false-alarm rate* (FA), the probability of responding yes on noise trials, describe performance on a *yes/no* task.

In *yes/no* tasks, and therefore in the present study as well, *d'* and c are calculated, according to Macmillan [71], in the following way:

$$d' = \Phi^{-1}(HR) - \Phi^{-1}(FA) \tag{1}$$

$$c = -\frac{\Phi^{-1}(HR) + \Phi^{-1}(FA)}{2} \tag{2}$$

Thus, *d'* is found by subtracting the z score that corresponds to the false-alarm rate ($\Phi^{-1}(FA)$) from the z score that corresponds to the hit rate ($\Phi^{-1}(HR)$). c, on the other hand, is found by averaging the z score that corresponds to the hit rate and the z score that corresponds to the false-alarm rate, then multiplying the result by negative one.

Regardless of the approach used for the Φ^{-1} function, problems arises when the hit or false-alarm rate equals 0, because then the corresponding z score is $-\infty$. Similarly, a hit or false-alarm rate of 1 corresponds to a z score of $+\infty$. These extreme values are particularly likely to arise when participants adopt extremely liberal or conservative criteria, meaning that they will answer only *yes* (liberal) or only *no* (conservative) to every trial.

Macmillan & Kaplan [72] suggested one possible approach to solve this problem. This approach involves adjusting only the extreme rates themselves. False alarm rates of 0 are replaced with $\frac{1}{2N}$, and false alarm rates of 1 are replaced with $1 - \frac{1}{2N}$, where N is the number of trials. In the current study, this approach was applied because these corrections were needed to do for the tests 4-6 in block 2 (Spoken words) where the number of trials ($N = 51$) was considered to be large enough to not affect the calculations of *c* and *d'* significantly. Similar replacements were not needed to be done for block 1 (Sounds) since no extreme values were obtained.

3.4.3 Analysing data with repeated measures ANOVA

Repeated measures ANOVA was considered as a data analysis tool for the current study because it has been used in other similar studies (see e.g. [27, 28, 30]). Furthermore, as Coolican [73] states, the rationale for repeated measures ANOVA is that people in the different conditions are not different people, and all the variations (between participants and within conditions) among all the scores need to be considered. This holds in the present study also since all the participants have gone through all the six different tests, and thus there is data for each participant in each test.

Data assumptions for repeated measures ANOVA

For the use of repeated measures ANOVA some assumptions needed to be first met. As with other parametric tests, it should be able to assume that the dependent variable is at interval level and is normally distributed. In addition, it should be able to assume *sphericity* of the dependent variable among the variables that make up the levels of each factor. The dependent variables in the current study are the parameters *d'*, *hit rate*, *c*, and *mean reaction time*. The independent variables and the factors are the blocks (Sounds and Spoken words), and the conditions (unisensory, with pictures, and with text) make up the levels of each factor. (For a reference for this topic see e.g. [73, 74].) In the following, the assumptions are presented.

1. Dependent variables at interval level

   Interval level data have equal amounts for equal measures on the scale; the dependent variable is measured at the continuous level. The data for the current study has been collected according to how participants performed in different tests, and the tests have been identical for every participant. In the tests participants answered by pressing one of the two keys on the keyboard; the other key expressing the answer *yes*, meaning "I heard this sound/word in the encoding part" and the other key expressing the answer *no*, "I did not hear this sound/word in the encoding part". If the participant answered and remembered correctly whether they had heard the sound/spoken word in the encoding part, the software reported "1", and if the answer was incorrect (e.g. answering "yes" although the sound/spoken word was a new one (a noise) and the correct answer would have been "no") the software reported "0". Therefore, a binary data set was collated where *ones* represented correct answers and *zeros* false answers. The dependent variable *hit rate* was extracted by calculating the percentage of correct answers (ones) in retrieving correctly the sounds/spoken words that had been presented in the encoding phase. Thus, the values of hit rate varied between 0% and 100% thus representing an interval level data.

The same can be said from *d'* and *c* since these parameters are derived from *hit rates* and *false alarm rates* that both varies between 0% and 100% (see Eq. 1 and Eq. 2). *d'* can get values greater than zeros but not more than 4.00, and the values of *c* can vary between -1.00 and 1.00. Additionally, mean reaction time meets this assumption because the values for mean reaction time were recollected from the participants' answers by calculating how much time every participant spent answering at each sound/spoken word. The reaction time was measured in milliseconds and the values varied between 700 ms and 4000 ms.

As a conclusion, it can be stated that all the dependent variables (*d'*, *hit rate*, *c*, *mean reaction time*) used in this data analysis with ANOVA, are at interval level and thus continuous. This first assumption is therefore met.

2. Approximately normally distributed data

   Since ANOVA is a parametric test it requires that the data population from which samples have been drawn is a normal distribution. This is because it is needed to be able to make fair assumptions about the nature of underlying sampling distributions. Whether the data was normally distributed or not, it was checked by drawing a Cumulative Distribution Function (CDF) from the measured data and by comparing that curve with the normally distributed predicted data.

   The results from the normality checks with CDF are presented in the appendix A in the Fig. A1, A2, A3 and A4. In the appendix A, there are all the results: for each dependent variable (*d'*, *hit rate*, *c*, *mean reaction time*) in each level i.e. in each condition (sounds-only, sounds with pictures, sounds with text, spoken words-only, spoken words with pictures, spoken words with text). The blue curve refers to the measured data and the red curve to the predicted data (normal distribution). It can be noticed, that all measured data reasonably follow the predicted data, and normality can be assumed in every test.

   The double check of the normality was done according to a rule of thumb [73] that normality can be checked by looking at the differences between the mean and the median; if they are far apart, a half a standard deviation, then there is a lot of skew and it cannot be said that the data is normally distributed. The double check (see Tab. A1) strengthened the assumption of normality, and thus it can be stated that all data approximates closely to the normal distribution and thus supports the use of ANOVA in the data analysis in this study.

3. Sphericity

   Sphericity of the dependent variables (*d'*, *hit rate*, *c*, *mean reaction time*) that make up the levels (unisensory, with picture, with text) of each factor (sounds and spoken words) should be able to assume. If the sphericity assumption is not valid, then the F test becomes too liberal which means that the test might give Type II error and thus might miss real effects. Type II error occurs when the true null hypothesis is rejected incorrectly, which then leads to not detecting an effect that actually is in the data. Sphericity occurs if:

   (a) there is homogeneity of variance among the level variables

   (b) the variances of the differences between levels are similar to one another.

   Homogeneity of variance requires that the variances in the two populations are equal and this is done by showing that the two sample variances are not significantly different. The null hypothesis assumes that both samples are drawn from similar distributions, thus the significance value for this should be greater than 0.05. Nevertheless, this assumption is met in this design already because the design is related (repeated measures) and the sizes of different conditions are all the same (N=42).

   In order to minimize the sphericity problem, Greenhouse-Geisser and Huyhn-Feldt indexes of deviation to sphericity can be used to correct the number of degrees of freedom of the F distribution. Greenhouse-Geisser and Huyhn-Feldt corrections (ϵ_{GG} and ϵ_{HF}) give values between 0 and 1 where 1 represent perfect sphericity. These epsilon values were calculated for each dependent variable for two-way and one-way ANOVAs, and the results are gathered in Tab. 1 and Tab. 2.

Table 1: Two-way ANOVA check: ϵ_{GG} and ϵ_{HF} values for each parameter.

d'		HR		c		RT	
ϵ_{GG}	ϵ_{HF}	ϵ_{GG}	ϵ_{HF}	ϵ_{GG}	ϵ_{HF}	ϵ_{GG}	ϵ_{HF}
0.9945	1.0450	0.9705	1.0180	0.8121	0.8408	0.9886	1.0384

   When looking at the values in Tab. 1, it can be noticed that both ϵ values for parameter c are clearly under 0.9. Therefore, c needs a new p-value that will be calculated from the ϵ_{HF}. This new p-value is calculated in the next section (Results Ch. 4). Similarly, c values in Tab. 2, especially in the data from Spoken words (Words), require a new p-value. Again the new p-value is gotten by using ϵ_{HF} and this the new p-value is presented in the next section (Results Ch. 4). All the other values are sufficiently close to 1 that it can be assumed approximately sphericity, and new p-values are not needed to correct.

Table 2: One-way ANOVA check: ϵ_{GG} and ϵ_{HF} values for each parameter in each block (sounds and words).

	d’		HR		c		RT	
	ϵ_{GG}	ϵ_{HF}	ϵ_{GG}	ϵ_{HF}	ϵ_{GG}	ϵ_{HF}	ϵ_{GG}	ϵ_{HF}
Sounds	0.9397	0.9833	0.9972	1.0482	0.9354	0.9785	0.9828	1.0319
Words	0.9812	1.0301	0.9636	1.0103	0.7543	0.7769	0.9471	0.9916

As it has been demonstrated above, all the data meet the repeated measures ANOVA assumptions, only two of the data require a corrected and new p-value. Thus, it has been shown that the use of repeated measures ANOVA is justified in the current study.

3.4.4 Data analysis workflow

The responses from the tests 1-6 were imported into Excel where the data analysis was executed as well. Two-way and one-way ANOVAs were applied first to test whether there is a significant interaction in the data and if yes, the pairwise comparisons were done with post-hoc t-tests.

In ANOVA, Block and Condition worked as repeated-measures factors. The Block had two levels: Sounds and Spoken words. The Condition, on the other hand, had three levels: unisensory (auditory), auditory with picture and auditory with text. The analysis was carried out for values of *d’*, for *hit rates*, for values of *c*, and for *mean reaction times*. These parameters were the dependent variables of ANOVA as described previously.

For each dependent variable, a two-way repeated-measures ANOVA (2x3) was conducted first, and based on the effect of interaction, data analysis was proceeded with one-way repeated measures ANOVA (1x3). Further, based on the one-way repeated-measures ANOVA, pairwise comparisons were made using post-hoc t-tests for means. Bonferroni correction ($= \alpha/$number of comparisons $= 0.05/2 = 0.025$) was applied in paired t-tests. Bonferroni correction produced the value of two in the denominator because the comparisons were made only between two Conditions: multisensory (with pictures or with text) vs. unisensory conditions. The original degrees of freedom and corrected p-values are reported with F values. The significance level was set to $\alpha = 0.05$.

4 Results

Values for *d'*, *hit rate* (the percentages of correct responses), values for *c* and *mean reaction times* (RT) for hit rates (hit) and for false alarms (false), and their standard deviations (SD) are presented in Tab. 3.

Table 3: Means and standard deviations for different parameters for both Experiments, Sounds and Spoken words (words), in different Conditions (unisensory, with pictures and with text).

Test	d'	SD	Hit rate (%)	SD	c	SD	RT hit (ms)	SD	RT false (ms)	SD
sounds	0.40	0.24	53.1	11.8	0.10	0.28	1588	487	1508	520
sounds& pictures	0.71	0.44	54.8	12.5	0.19	0.27	1831	601	1951	838
sounds& text	0.71	0.43	57.5	14.5	0.13	0.37	1739	655	1824	675
words	1.37	0.64	60.6	18.0	0.36	0.44	1396	362	1420	270
words& pictures	2.37	0.70	75.8	16.3	0.35	0.58	1333	293	1370	244
words& text	1.71	0.71	58.9	18.1	0.58	0.46	1391	399	1391	363

4.1 d' showed enhanced memory performance in audiovisual conditions

The recognition memory performance as *d'* for different Conditions (unisensory, with pictures, with text) divided according to the Block (Sounds = orange colour, Spoken words = blue colour) is shown in Fig. 6.

Two-way repeated measures ANOVA for *d'* revealed strongly significant main effects for Block [F(1,246)=217.28, p<.0001], for Condition [F(2,246)=50.84, p<.0001] and for Interaction [F(2,246)=18.06, p<.0001], suggesting that there were differences between the Conditions (unisensory, with pictures, with text) and Block (sounds, spoken words). The significant result from Block suggested that it mattered whether participants needed to encode sounds or spoken words; there was significant difference between these two. The significant result from Condition, on the other hand, represented the differences between the three conditions, it mattered whether the auditory stimulus was presented alone, with pictures or with text. With the significant result from Interaction it was continued with further analysis to state the true main effects in the data.

In one-way repeated-measures ANOVA for *d'*, the main effect was significant for both, for Conditions of Sounds [F(2,82)=10.94, p<.0001] and for Conditions of Spoken words [F(82,2)=59.18, p<.0001]. This implied that Conditions inside Block varied. To find the statistically significant differences further analysis was carried with paired t-tests.

Paired t-tests for Conditions of Sounds revealed that sounds with pictures [t(41)= -4.54, p<.0001] and sounds with text [t(41)= -4.07, p<.0001] were significantly better remembered than sounds-only. However, between sounds with pictures and sounds with text there was no significant difference [t(41)= -0.11, p=0.91]. Paired t-test for Conditions of Spoken words revealed similar type of results since spoken words with pictures [t(41)= -9.90, p<.0001] and spoken words with text [t(41)= -3.31, p=.0019] were significantly remembered better than spoken words alone. In addition, spoken words with pictures were remembered better than spoken words with text [t(41)=5.86, p<.0001]. These results are shown in Fig. 6.

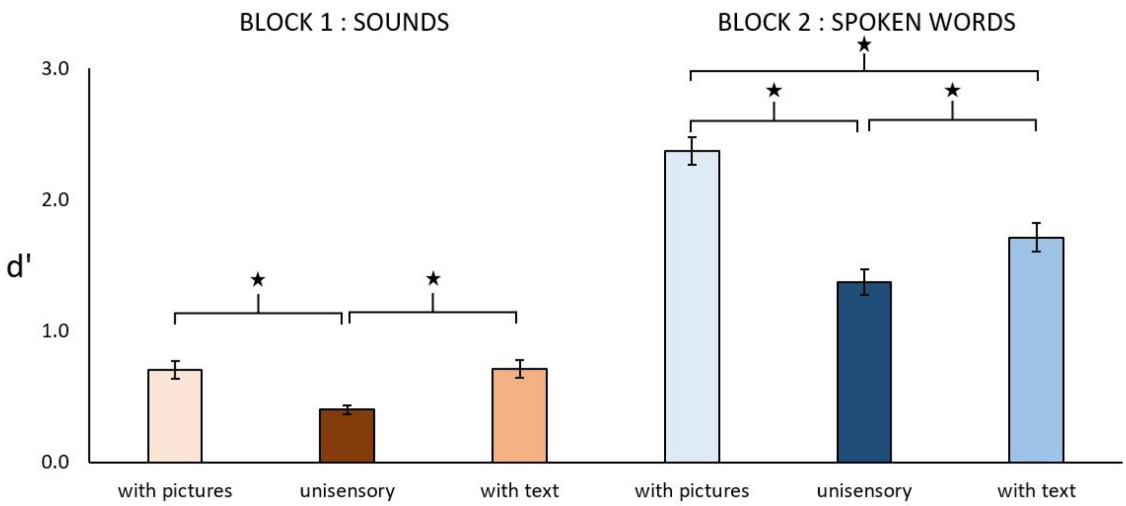


Figure 6: The recognition memory performance in two blocks presented as values of *d'* for different Conditions: with pictures (light colour), unisensory (dark colour), with text (intermediate colour) of Sounds (orange colours) and Spoken words (blue colours).

4.2 Hit rate showed different memory performance than d'

In two-way repeated-measures ANOVA for hit rate, the main effects of Block [F(1,246)=19.69, p<.0001] and Condition [F(2,246)=17.33, p<.0001] were strongly significant. Importantly, the Interaction between Block and Condition was also strongly significant [F(2,246)=24.16, p<.0001]. This arose because different Conditions showed different effects of Block.

Now, interestingly, in one-way repeated-measures ANOVA for hit rate, the main effect of Condition for Sounds was not significant [F(2,82)=2.27, p=0.11], but on the other hand, it was strongly significant for Spoken words [F(2,82)=37.82, p<.0001]. This result implied that there was no differences between the Conditions in Block of Sounds but, on the contrary, there was at least one difference between the Conditions of Spoken words. To find the difference or the differences, a paired t-test was conducted.

According to the results from the paired t-test for Conditions of Spoken Words, spoken words with pictures was remembered better than spoken words alone (unisensory) [t(41)= -6.99, p<.0001] and better than spoken words with text [t(41)=7.36, p<.0001]. However, there was no significant difference between spoken words with text and spoken words alone. Paired t-test for Hit rate of Sounds was not conducted because the main effect in one-way repeated-measures ANOVA was not significant and thus implying that there was no difference between any of the Conditions of Sounds. These results are shown in Fig. 7.

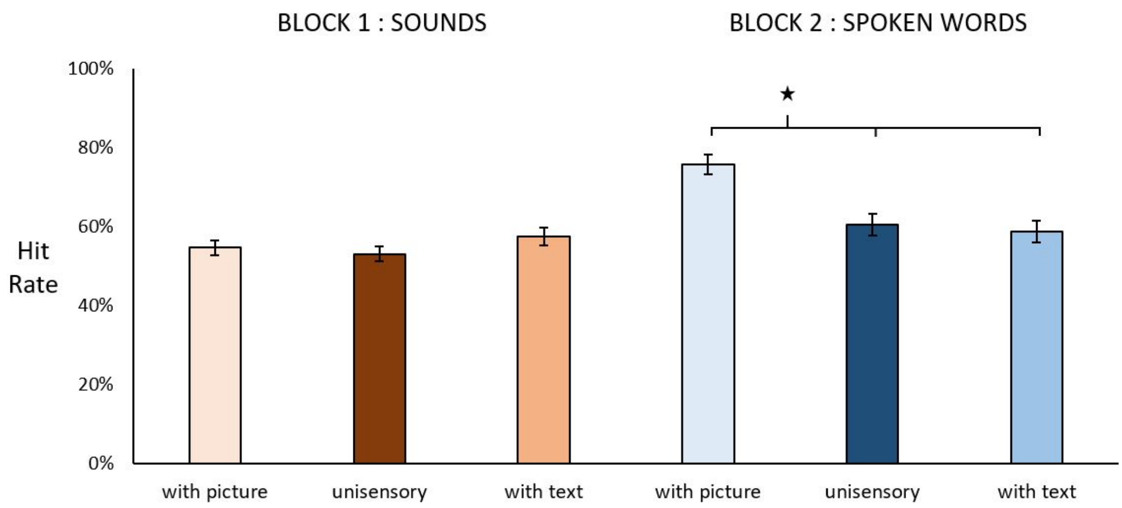


Figure 7: The hit rate presented for both Blocks, Sounds (orange colours) and Spoken words (blue colours), in three different Conditions; with pictures (light colour), unisensory (dark colour), with text (intermediate colour).

4.3 Criterion *c* varied significantly

As sphericity tests in Ch. 3.4.3 showed, the ANOVA values for *c* were needed to be corrected according to the ϵ_{HF} values (see Tab. 1 and Tab. 2). Two-way ANOVA for *c* therefore showed that the main effects of Block [F(1,246)=28.00, p<.0001] and Interaction [F(2,246)=5.64, p=0.005] were significant but the main effects of Conditions (F(1.68,206.84)=3.94, p=0.051] were not significant. This implied that *c* varied significantly depending on the Block (Sounds or Spoken words), and therefore the results gained from hit rate could not be trusted. This was the reason why signal detection theory was applied, and thus the trustworthy results of this study were obtained with the results from *d'* and not with hit rate.

Since Interaction was significant, further analysis with one-way ANOVA was carried for Blocks. The results of this analysis showed a significant main effect for Conditions of Spoken words [F(1.55,63.71)=5.55, p=.0216] but not for Conditions of Sounds [F(1.96,80.24)=2.50, p=.1175]. Since there were some differences in *c* depending on Conditions, it meant there was significant fluctuation, and therefore the comparisons between different Conditions of hit rate were not justified. Thus again, *d'* was used because it considered the fluctuation of *c* and therefore made the results more comparable.

The results from one-way repeated-measures ANOVA for *c* showed significant difference in Conditions of Spoken words, and to identify the differences, paired t-tests were conducted. According to pairwise comparisons for Conditions of Spoken words, *c* was higher for spoken words with text than for the unisensory condition [t(41)= -4.19, p=.00014] or for spoken words with pictures [t(41)= -2.56, p=.014]. *c* between spoken words-only, and spoken words with pictures did not show any significant difference [t(41)=.20, p=.84]. The results from *c* are shown in the Fig. 8.

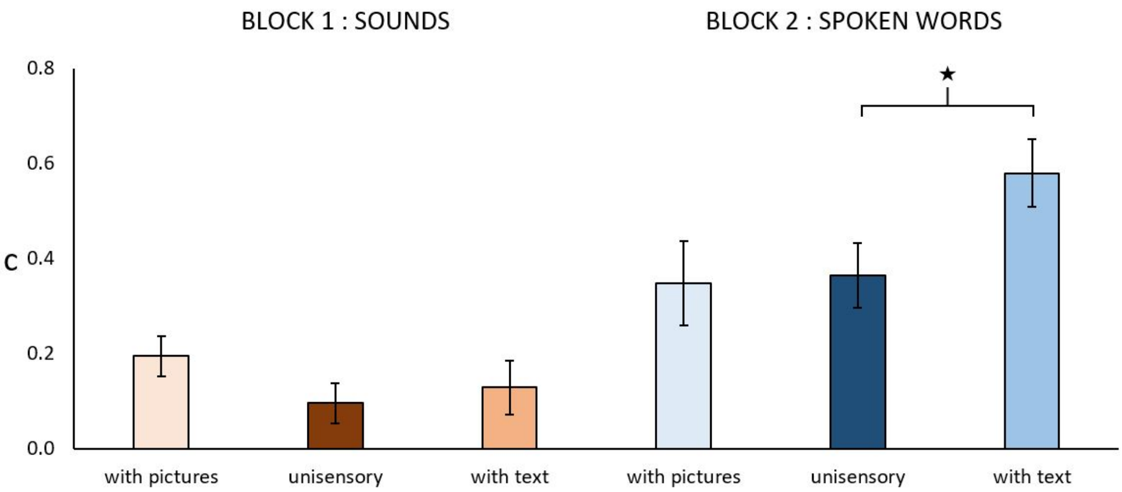


Figure 8: The criterion *c* presented for both Blocks, Sounds (orange colours) and Spoken words (blue colours), in three different Conditions; with pictures (light colour), unisensory (dark colour), and with text (intermediate colour).

4.4 Response times were longer for Sounds than for Spoken words

In the two-way repeated measures ANOVA for mean reaction times, the analysis revealed interesting results. Significant main effect was found for Block [F(1,246)=32.20, p<.0001], and for Interaction [F(2,246)=6.22, p=0.003] but not for Condition [F(2,246) =2.52, p=0.087]. This implied that Sounds and Spoken words showed different trends. Since Interaction was significant, further analysis was conducted.

The repeated measures ANOVA for Stimulus showed a significant main effect for Condition of Sounds [F(2,82)=6.53, p=.00234] but not for Condition of Spoken words [F(2,82)=0.88, p=0.42], meaning that only in Sounds the Conditions varied significantly and showed differences. Therefore further analysis with paired t-tests for Sounds was carried. These pairwise comparisons showed that participants responded quicker to sounds-only than to sounds with pictures [t(41)= -3.79, p=.00049] or to sounds with text [t(41)= -2.25, p=.03]. This result was interesting because it referred that it took longer time for participants to respond in retrieval phase when in encoding phase they had multisensory stimuli with sounds. The results from mean reaction times are shown in Fig. 9.

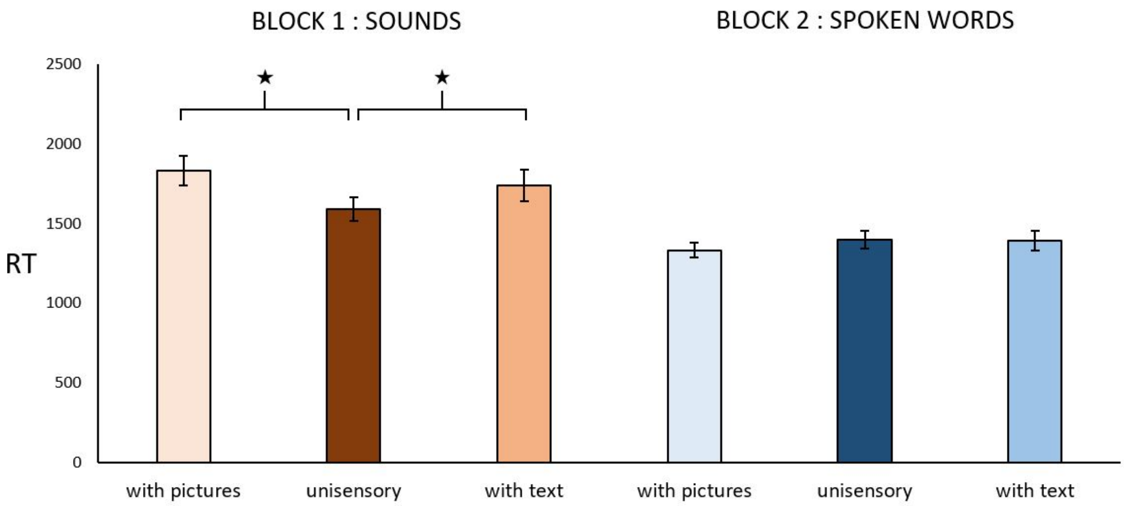


Figure 9: The mean reaction times presented for both Blocks, Sounds (orange colours) and Spoken words (blue colours), in three different Conditions; with pictures (light colour), unisensory (dark colour), and with text (intermediate colour).

5 Discussion

5.1 Summary of multisensory findings

The primary goal of this study was to determine whether older adults benefit from visually presented semantically congruent stimuli in the improvement of auditory memory. It was hypothesized that in the encoding phase audiovisually presented semantically congruent stimuli would lead to enhanced auditory recognition memory performance in the retrieval phase. Additionally, spoken words were hypothesized to be generally remembered better than sounds. It was also hypothesized that there would be no effect in response times.

The present study shows that older adults can benefit from audiovisual semantic congruency during memory encoding, which is in line with previous studies conducted with young adults [22–29] and with children [30]. The recognition memory performance expressed as *d'* was enhanced when auditory stimuli were paired with semantically congruent visual stimuli during encoding compared to auditory-only unisensory condition. Older adults' auditory memory performance was better for all multisensory (audiovisual) conditions compared to unisensory conditions in both blocks, in Sounds and in Spoken words. The greatest benefit was achieved when spoken words were presented together with pictures. Memory performance in Sounds in general was surprisingly weak. Response times revealed, against the expected, significant differences: response times for sounds were longer than for spoken words but response times for sounds-only condition were faster than for multisensory conditions. Criterion c varied significantly too resulting in unreliable results for *hit rate*.

Taken together, these results provide empirical evidence that audiovisual semantic congruency facilitates recognition memory performance in older adults. The present study provides the first evidence, based on current information, that older adults can benefit from multisensory memory cues. Luo et al. [32] did not find this effect in older adults, maybe because their participants were instructed to recall the visually presented written words that were originally presented together with natural sounds. As it has been shown by previous studies conducted with young adults, auditory cues do not help that frequently to remember visual stimuli as visual cues help to remember auditory stimuli in the later recognition [24, 25, 27, 28]. Also, there is evidence that auditory memory is inferior to visual memory [31], that is, auditory cues do not help to remember visual cues because eyesight dominates. This then can explain the differences in the results between Luo's et al. study and the current study, since in Luo's et al. study participants were asked to remember visual cues that were originally presented together with auditory cues. However, in the current study it was reversed because the participants memorized auditory cues accompanied visual cues.

5.2 Sounds showed poorer performance compared to Spoken words

Despite the hypothesis that the memory performance for sounds would be poorer than for spoken words, that poor performance for Sounds was unexpected. Especially in sounds-only (unisensory condition) the results (d' = 0.40 and hit rate = 53.1 %) seem to refer to a nearly random response without any true memory trace (answering *yes* and *no* randomly). In other words, the participants could not differentiate between the old (stimulus presented in the encoding phase) and new (stimulus presented first time in the retrieval phase) stimuli. The participants' informally observed reactions during the experimental run in sounds-only test can give extra support for the results. While performing the tests, generally they showed greater difficulties with sounds and especially with the sounds-only test. In the experiment while trying to distinguish the old and the new items in the retrieval phase, they were shaking their heads and sighing with frustration, and after the Sounds block many of the participants, if not everyone, commented how difficult the task was and how they had almost no clue which sound was presented before and which not. Additionally, if the criterion c is 0 then the participant's criterion is *neutral* showing no decision bias [70], and for sounds-only this criterion was 0.10 and thus being close to 0, which supports the idea that participants were almost randomly choosing between *yes* and *no* without really remembering in which way it truly occurred.

If, on the other hand, this is compared to the spoken words experiment, the results for c for Spoken words experiment, show different trend. The condition of spoken words with text has c of 0.58. The criterion of clearly over 0 shows bias toward responding *no* [70], which in this case implies that older adults did not answer *yes* before they were very sure about knowing and remembering the answer correctly, rather they chose to answer that they do not remember whether the word was presented before or not. The similar trend, rather clearly a weaker response bias, is shown with spoken words-only (c = 0.36) and spoken words with pictures (c = 0.35) conditions. This was also seen informally during the experiment run where participants spoke aloud whether they knew the answer or not and the answer was mostly *no*. However, the *yes* answers were very positive implying that then they really knew the answer and remembered it. Informally observing, similar reaction rarely happened with the sound experiment, which can give more insight in that the spoken words leave stronger memory traces into participant's recognition memory than does sounds based upon the current study design.

Why then do older adults find a greater difficulty identifying between the old and new sounds in the retrieval phase in the Sounds block? Possible explanations can be related to information processing speed and hearing range. Reduction in the speed of cognitive processing can partly explain the difference between the memory performance in Sounds and Spoken words, since the slowdown of general information processing has been noticed with ageing (see e.g. [39]). In this experiment, sounds were presented for a short period of time (400 ms/stimulus), which might have been too fast for older adults for efficient information processing. Whereas in Spoken words the stimuli were easily understandable and recognizable one or two syllable words and thus probably quicker to process and retrieve than sounds. According to a major hypothesis (see e.g. [75]), spoken words might have been recognized in the context of other words in memory, and thus processed quicker and more efficiently. This then might have affected the accuracy and speed of recognizing the words.

Another possible explanation can be the sensory-level changes in normal ageing. With age, the sensitivity to higher frequencies gradually lowers, that is, the hearing range narrows (see e.g. [76]). However, the narrowing hearing range does not alone lead to difficulties in hearing since people with normal hearing range might still notice difficulties in certain listening conditions, particularly when noise is present it might be the case of encoding supra-threshold sounds [77]. Thus, it could be explained that older adults might not be able to really hear the sounds as clearly as young adults hear them, which would be due to narrowing hearing range and weakened encoding of supra-threshold sounds. That is, it can suggest that older adults might not be able to differentiate between all the nuances the sounds carried, and thus failed recognizing sounds close to each other (e.g. recognizing a bell from a phone).

In the encoding of sounds, during the experiment, the environment might have affected many of the participants as they were studied at their homes in silent rooms. However, it was impossible to make the room totally silent and therefore some external noise could have been intermingled with the sounds of the tests causing confusion and poor encoding. External noise, on the other hand, does not appear to affect spoken words with the same extent and this can be due to the understandability and recognizability of these familiar and widely used spoken words (see List of Stimuli in Appendix B).

5.3 Why does semantic congruency help older adults?

Clearly older adults benefit from semantic congruency since in this study only semantically congruent audiovisual stimuli were used to test the recognition memory, and the overall memory performance was better every time for audiovisual condition than for auditory unisensory condition. The benefits from semantic congruency have been identified widely in the audiovisual memory studies conducted with young adults [22–29, 78] and with children [30]. Also, non-semantic congruency effect have been noticed in audiovisual integration studies when young and older adults have been compared [13–15]. Thus, there is strong evidence that people of all ages benefit from congruency and this study is in line with that evidence. However, it could be evaluated to a greater depth why congruency helps older adults.

In the current study, the congruency effect is gained when auditory stimuli is originally presented together with visual stimuli. This implies a strong audiovisual integration. Previous research has shown that older adults tend to lean toward visual cues rather than auditory cues [16–18]. In one study, it was suggested that the enhanced visual influence for older adults could have been associated with an age-related delay in auditory processing [18], and other similar studies suggested the same that visual influence for older adults is more due to perceptual rather than cognitive processing [19, 60]. Thus, older adults' tendency on seeking a support from visual cues can have affected the congruency effect in the current study too, although no additional support was given in the retrieval phase. By good perception of the visual stimuli (pictures or written words) in the encoding phase, stronger memory traces can have been built when the originally audiovisually paired auditory stimuli (sounds or spoken words) can have been able to recall with better accuracy later in the test, which can refer to the congruency effect found in the current study.

A large visual influence has also been seen among older adults with hearing loss, and the greater hearing loss the larger benefit from visual support [63]. Hence, older adults can compensate from perceptual losses by leaning more on the modality with good and clear signal clarity than on the impaired modality or on the modality with poor signal clarity [16, 17]. A neuroimaging study has also shown that the facilitation of neural responses occurs earlier in and to a greater extent in older adults than in younger adults although the overall performance would have been very similar between the age groups [62]. The same study deducted that older adults benefit more from visual cues than young adults, to perhaps compensate for sensory ageing, which can support the result of the obtained congruency effect in the current study too.

Compensatory benefits for declines in sensory processing have been previously proposed as a possible explanation of a congruency effect for older adults [13], which appears to be in line with the principle of multisensory integration that weak stimuli are more likely to be integrated [10]. Thus, it can be suggested that older adults benefit more from receiving redundant information across multiple sensory channels. The Principle of Inverse Effectiveness (PoIE) proposes similarities. Previously PoIE has been tested in speech perception studies because PoIE predicts that older adults will show enhanced integration during audiovisual speech recognition relative to young adults [61]. The idea is, that reduced sensitivity in the individual sensory systems combined with age-related alterations in cognitive processing increases the relative magnitude of multisensory enhancements [79]. This then leads to a greater importance of multisensory integrations during old age, which helps to counteract the consequences of unisensory deterioration. Thus, this can be why, in the current study, memory performances in all multisensory conditions are greater than in unisensory conditions for older adults.

On the other hand, Heikkilä et al. [27] were the first ones to show that there is a congruency effect also in a longer-term memory performance and not only in very short-term memory tasks (see e.g. Murray and co-workers [22, 23, 25]). Heikkilä et al. thus proposed that congruent multisensory stimuli may receive more effective encoding than unisensory stimuli [27]. This is in line with the results obtained in the current study. Murray et al. have shown that for multisensory stimuli, a larger set of processing structures are involved and thus multisensory structures are also active with visual and auditory structures [22]. Therefore Heikkilä et al. concluded that, congruent stimuli lead to a more extensive and enhanced activation than a single stimulus resulting in a richer memory trace. This memory trace can be then activated by a later presentation of a unisensory stimulus, which can lead to a better recognition memory performance, as has been shown in the current study.

Very recently Heikkilä et al. [28] made an interesting proposal for the congruency effect in their similar study to the current study, except their study was conducted with young adults. They considered that the congruency effect could be more "volatile" for verbal than for non-verbal material because congruent spoken words did not facilitate the recognition memory for written words, or vice versa. In Heikkilä's et al. previous studies they have obtained partly dissonant results and partly coherent results. In their studies, they have clearly shown that pictures facilitate the later memory retrieval of sounds [27, 28, 30] and spoken words [28–30]. However, the facilitation of written words for later memory retrieval with sounds or spoken words have not been as coherent. For example, neither school-aged children [30] nor young adults [28] have benefited from written words when recalling spoken words, although Heikkilä et al. obtained a benefit previously with young adults [27]. Written words have, on the other hand, facilitated the later memory retrieval of sounds in young adults [28].

The current study, on the other hand, shows that older adults can benefit from both, written words and pictures, that have been presented together with sounds or spoken words in the encoding. This finding for its part supports therefore the proposal of Heikkilä et al. [28] that the congruency effect can be more volatile for verbal and for non-verbal material because pictures have shown to offer congruency effect benefits every time whereas written words not. It seems that there might be some sort of age effect because children show weak verbal effect [30], young adults may have the verbal effect [27, 28], and older adults do have the verbal effect (the current study).

5.4 Older adults seem to show a great picture-facilitation effect with verbal material

The memory performance in spoken words with pictures condition were clearly better in the Spoken words block compared to any other condition in the Sounds block or in the Spoken words block. It was expected that pictures will facilitate later recognition memory performance, as the picture-facilitation effect has been discovered by Heikkilä and co-workers for young adults [27, 28] and for school-aged children [30], but this great facilitation was unexpected.

In Heikkilä's et al. very recent studies [28, 29] a picture superiority effect is suggested in recognition memory research. It is indicated that pictorial information might be more effectively processed than other forms of information, that is, congruent pictures facilitate memory encoding more than other kinds of stimuli due to its more conceptual nature. In the current study, the picture superiority effect is seen especially in Spoken words block because the "with pictures" condition showed the best memory performance. However, the same pattern cannot be seen in Sounds block because the "with pictures" condition is not remembered any better than the "with text" condition although both conditions showed better recognition memory performance than unisensory auditory condition.

Heikkilä et al. [28] discussed their results compared to another model concerning working memory by Baddeley (see his review [80]). In the current study conducted with older adults, similar results were gained, as in Heikkilä's et al. study conducted with young adults [28]. Also in the current study, it was found that pictures improved the memory of sounds and spoken words. The working memory model, then, suggests that working memory consists of a central executive system and its slave systems that are a phonological loop, a visuospatial sketchpad and an episodic buffer. The phonological loop handle phonological and acoustic information, and auditory verbal information is assumed to enter it automatically but visually presented words are considered to transform into a phonological code through reading and silent articulation. Thus, the model suggests that all verbal information is processed in this phonological system despite the presentation modality. Pictures, on the other hand, are processed by different system, in visuospatial sketchpad. This lead Heikkilä et al. [28] into a proposal that picture-sound/word congruent audiovisual stimulus pairs might be processed more efficiently than other stimulus pairs or unisensory stimuli because pictures are processed in different systems compared to sounds/words. They concluded that more elaborate memory trace can be achieved by processing semantically congruent information in different systems during encoding.

The final possible explanation behind the picture facilitation effect is offered by Diaconescu et al. [81] who proposed that visual dominance effect gets stronger with age. This was concluded because in their study older adults showed a more pronounced gain in performance during multisensory trials compared to auditory unisensory ones. Additionally, older adults responded faster to complex sounds when they were accompanied by visual stimuli. Diaconescu et al. discussed that the visual dominance may have been more pronounced due to the nature of their semantic classification. Although in the current study older adults responded slower to sounds when they were originally presented together with a congruent visual stimulus than when no additional information was given, visual dominance effect can still explain the picture facilitation effect. Rather, it could be proposed that visual dominance effect in older adults is especially observed when pictures are presented concurrently with verbal material. On the other hand, visual dominance effect can be seen with written words too, and not only with pictures. Thus, it seems that any supportive visual material can help older adults.

Taken together, there are no simple explanation why older adults show great picture facilitation effect with verbal material. It is possible that the facilitation effect is observed because pictures and spoken words are processed in different systems, in accordance with the Baddeley's working memory model, and therefore semantically congruent information processed in different systems during encoding might lead to a more elaborate memory trace [28]. It is possible that the picture superiority effect and the visual dominance effect play a role in this facilitation effect.

5.5 Explaining findings from the point of view of the CSTM model

Potter's Conceptual Short Term Memory (CSTM) model [82, 83] was first presented in audiovisual recognition memory research by Heikkilä et al. 2015 [27]. They explained that the CSTM model can provide a theoretical framework to explain how encoding is enhanced for congruent stimuli. Before Heikkilä et al., Chen & Spence already suggested that audiovisual semantic interactions could occur in this CSTM [84]. Thus, according to them, semantic representations of multisensory stimuli could be rapidly accessed and temporarily retained in order to form a coherent multisensory object representation. Originally Potter has described CSTM to be a mental buffer and processor that rapidly integrates new perceptual information with information held in the long-term memory, and creates new representations that can be encoded into the long-term memory [82, 83].

CSTM proposes, according to Potter [83], that when perceiving a meaningful visual stimulus (a word, a picture or an object) it is rapidly identified and this then activates associated information from long term memory. Between these active concepts, new links are formed in CSTM and these links are shaped by perceptual information and current goals. The resulting structure is conscious and representing the understanding of the perceived visual material. The structure is consolidated into long term memory, and information that has not been incorporated into these structures will be rapidly forgotten.

As Heikkilä et al. indicated, CSTM has generally been thought to operate only with visual material since there has not been any studies where auditory stimuli would have been applied. However, both Heikkilä et al. and Chen & Spence have proposed that CSTM is a structure that can utilize visual as well as auditory information for providing coherent information that enhances the encoding [27, 84], which would be in line with Potter's original suggestion [82]. Therefore, in the current study, it is argued that when a simultaneous congruent auditory information is presented with a semantically congruent visual information, the visual information provides coherent information that enhances the encoding of the representation of the auditory information into CSTM.

CSTM can also generally explain why auditory recognition memory performance was better in audiovisual conditions than in unisensory conditions in the current study. This extended CSTM can explain how congruent audiovisual stimuli facilitate the auditory recognition memory and thus, why e.g. semantically congruent pictures facilitated the later recognition of spoken words. When the participants recalled these spoken words that were initially presented with congruent pictures in the encoding phase, these congruent pictures then enhanced the encoding of spoken words into CSTM by providing coherent information, in line with the Heikkilä's et al. proposition [27]. Similarly happens when written words are presented together with sounds or spoken words since written words are meaningful visual stimuli too.

CSTM can offer an explanation model for the poor memory performance in Sounds compared to Spoken words. As discussed earlier, sounds were presented for a too short period of time that older adults could not process and identify the sounds at that required speed. Therefore, the auditory information (sounds) stays less meaningful and cannot be identified rapidly at a conceptual level. This then leads, according to the CSTM model, that no new links are formed between the active concepts in the CSTM, and the information will not be incorporated into these structures but are rapidly forgotten. This, on the other hand can be related to the picture superiority effect. It requires that the concurrent other stimuli are well identified too, as with spoken words, for it to occur with success in older adults. However, with sounds this might not happen when the picture superiority effect stays weaker in Sounds than in Spoken words.

Taken together, Potter's CSTM model can offer interesting explanations for the results obtained in the current study. CSTM can explain generally why the memory performance in audiovisual conditions is better than in unisensory auditory conditions. Additionally, it can allow a theory for why the overall performance in Sounds failed to match the overall performance of Spoken words. Picture facilitation effect can, for example, occur also because both pictures and spoken words carry a lot of conceptual information and can be rapidly identified and associated producing a conceptual structure that be held in the working memory or moved into the long-term memory, as suggested in the CSTM model. Future research could better test the functionality of CSTM as a theoretical framework for audiovisual gain in younger and in older adults.

5.6 How older adults compare to their younger counterparts

The closest study, where to compare the results gained in the current study, is the Heikkilä's et al. study from this year [28]. As discussed already, older adults benefit from written words in the encoding of spoken words, whereas young adults do not. This can possibly be explained as older adults put more emphasis on visual cues and young adults rather count on auditory cues [16, 18].

Generally, young adults appear to perform better than older adults, especially in Sounds block where the results from young adults are easily over 1.0 [28], but the results from their older counterparts are clearly under 1.0 (the current study). Young adults performing better than older adults have been previously recognized in multisensory perception studies where young and older adults have been compared [15, 21]. The difference in performance in Sounds between young and older adults can be due to older adults' lowered processing capacities which might have resulted in troubles of processing briefly presented stimuli [39], as discussed earlier.

Most interestingly, it seems that young adults benefit from audiovisual condition more than older adults in Sounds block, but older adults seem to benefit from audiovisual condition more than young adults in Spoken words block. This can suggest that older adults benefit from multisensory presentations when encoding spoken words more than young adults, which can have real life situation benefits since problems in speech perception are common among older adults. Other semantically congruent visual information can thus help older adults to resolve the phonetic and semantic information in unpredictable speech. Notable, there are already results gained in that direction. Maguinness et al. [21] observed that older adults' performance was enhanced more when non-meaningful sentences were accompanied with information in other modality compared to meaningful sentences with additional information in another modality.

5.7 Older adults require more response time when processing complicated sounds that were originally presented audiovisually

Contrary to what was predicted, mean reaction times showed significant difference in the current two-phased recognition memory study. However, the effect was seen only in Sounds. Surprisingly, in the multisensory conditions the participants answered more slowly than in the unisensory condition (sounds-only). Similar results have been seen in continuous recognition tasks were significantly slower reaction times have been produced for multisensory conditions [23], however, sometimes no effect between unisensory and multisensory condition have been obtained [24, 25]. On the other hand, mean reaction times have shown significance and explained multisensory enhancement only in simple detection task [81], in object recognition task [78], and in forced choice discrimination task [85]. Interestingly, in similar recognition memory tests as the present study, only Heikkilä et al. [27] have presented mean reaction times, but the results showed no statistical significance. Other studies [28, 30–32], have not even reported mean reaction times which is signalling that reaction times might have not shown any effect and thus not brought any importance to the findings and to the study.

Mean reaction times in Sounds (1719 ms) were significantly slower than in Spoken words (1374 ms). This again could be explained by the difficulty of differentiating between the sounds and correctly recognizing them. The presentation time in the current study was relatively short (400 ms) which must have made the recognition of the sounds more difficult than, for example, in the study by Cohen et al. where sounds were presented for five full seconds [31]. Thus, processing of sounds requires more effort, which is implied by the results from hit rate (close to 50% correct) and criterion c (close to 0, no decision bias), and it can be seen in slower reaction times than in Spoken words block.

On the other hand, this finding that multisensory condition actually slows the response times in Sounds, is consistent with previous studies where the continuous recognition task has been used [22, 23, 68]. However, these studies have asked participants to recall the visual, and the sound has been added in an audiovisual condition. In the current study participants always recalled the auditory, and the visual cues were only used for supporting and strengthening the encoding for later recognition.

Lehmann et al. [23] concluded that the slower reaction times on multisensory trials provides an indication that participants were unaware of the fact that only initial presentations included sounds. This, however, is not the case with the current study because participants knew which condition was happening at which time. However, one plausible explanation in this case would be the level of difficulty of recognizing the sounds combined with the visual support which adds more information for the processing and therefore slows it down. Therefore, it seems that the retrieval process in older adults slows down because in the test phase (during the retrieval) the stimuli are presented only as auditory and not as audiovisual.

5.8 Limitations

One clear limitation in this study are the circumstances where the study took place. The participants were examined in different locations depending on their own preferences. The places were in the universities' classrooms, laboratory room, or in participant's own house and there in the living room or in other closed room. This was done to enable better and easier participation to the study, to lower the threshold to participate. Due to the variation of the location, the background noise was not equal for every participant, additionally, it was impossible to fully block out any external noise or sound, although everything possible was done to ensure that there is the least amount of disturbance.

However, it is recognized that unwanted and uncontrolled background noise might have impacted upon the current study. Alain and Woods [86] have shown that background auditory stimulation (i.e. noise) presented during the performance of a visual task resulted in larger brain responses in older than in younger adults. This data indicated that older adults process more background information than young participants resulting in older adults making many more errors on a visual-discrimination task when background auditory noise is present. Therefore, there is a possibility that external noise and disturbance might have affected the results, although it does not seem very probable since participants did not change the room during the examination. However, the external noise might have influenced especially the recognition of sounds by causing interference since the results showed a poor performance in Sounds block. On the other hand, the external noise does not seem to have influenced the recognition of spoken words because the overall performance was good.

Another clear limitation is that cognition, memory or hearing thresholds were not tested for the participants. Surely they were asked whether they had hearing or sight difficulties or if they are diagnosed by a cognitive or a memory deficit. However, it is still not known, with certainty, whether they had normal hearing and normal cognitive functioning and thus it cannot be tested whether there would have been some correlation between memory functioning and the results, or between the hearing deficits and the results. Only visual acuity was tested and the participants, whose data was selected into the data analysis, had satisfactory vision.

5.9 Future prospects

The findings from this study opens new, interesting avenues for future research on the benefits of multisensory memory and learning in later life. It seems that the brain capacity of elders should be utilized more effectively, especially during working life. The foreseeable development of the age structure already supports this concept.

Generally, in the future studies it would be beneficial to examine how multisensory processing could help to maintain the good cognitive functioning in old age. Since for society, good cognitive processing of ageing population matters, as Docent Timo Pohjolainen expresses "There are two vital questions for political economy. How working population stays in working condition? How do elderly people stay self-sufficient for as long as possible in the home-like environment?" (translated from Finnish into English [42]). As a prerequisite for managing at home, mental agility is essential. It is much harder to compensate cognitive weakening than physical limitations. The ability for good cognitive processing even in old age is a goal worth pursuing.

One possible topic for future research could be a similar study to that described in this thesis but conducted with people with mild cognitive impairment or early-stage Alzheimer. It would be interesting to investigate which combination of multisensory information facilitates the most in these cognitively impaired people. With the finding of this type of study, it could be then examined and the information utilized in rehabilitation and if there are positive findings, and then how the findings could be used to aid such people.

Another possible future research topic could be the application of semantic multisensory information in learning new languages in later life. Fascinating findings can be revealed when investigating how older adults learn and remember better and more efficiently new words and meanings of a foreign language when the information is taught with semantically congruent multisensory information.

6 Conclusions

This thesis presented a multisensory memory study that investigated if semantically congruent multisensory experiences during encoding can improve memory performance in older adults. More precisely, semantically congruent visual stimuli during encoding do enhance later auditory recognition memory.

Older adults indeed performed better in conditions where semantically congruent visual stimuli were presented together with the auditory stimuli (audiovisual conditions) than in conditions where no additional information in another modality were given (auditory-only unisensory condition). Moreover, spoken words were remembered better than sounds as it was hypothesized, although the great difference between the performances was unexpected. There was a statistically significant difference in reaction times as the participants responded slower to audiovisual conditions than to auditory unisensory condition in Sounds block. Older adults gained a great benefit especially from semantically congruent pictures that were presented together with spoken words. This type of a picture facilitation effect was found for both Sounds block and Spoken words block but for Spoken words block it was particularly strong.

Findings gained in the present study follow the previous multisensory memory research conducted with young adults and with children. This is the first study to show that older adults can benefit from multisensory memory cues. These results show that semantically congruent multisensory experiences during encoding can improve memory performance in normal ageing. The main contribution of the current study is that it opens new, interesting avenues for future research on the benefits of multisensory memory and understanding better the learning in later life. The importance of this topic can be expected to grow in the near future due to the rapid change in population age distribution. The findings observed in the current study can be a great benefit in the future when planning new ways for maintaining good cognitive functioning at old age.

In the near future, the increase of the retired population results in a comparative decrease in the working population. Therefore, it is important to utilize the brain potential of elders, and maintain mental abilities to ensure autonomous life. Although the brains would have gone through many changes along the life, cognitive processing can still be enhanced by optimally using the remaining brain capacity. Despite the age, the learning ability of new skills remains, resulting in that the cognitive performance can be improved through practice.

References

[1] Suomen virallinen tilasto (SVT), "Väestörakenne 2016: verkkojulkaisu." http://www.stat.fi/til/vaerak/2016/vaerak_2016_2017-03-29_tie_001_fi.html, 2016. Accessed: 11.06.2017.

[2] Suomen virallinen tilasto (SVT), "Väestöennuste 2015-2065: verkkojulkaisu." http://www.stat.fi/til/vaenn/2015/vaenn_2015_2015-10-30_tie_001_fi.html, 2015. Accessed: 11.06.2017.

[3] United Nations, "World population ageing 2015." www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf, 2015. Accessed: 11.06.2017.

[4] National Institute for Healt and Welfare, "Memory disorders." www.thl.fi/en/web/chronic-diseases/memory-disorders, 2016. Accessed: 11.06.2017.

[5] Alzheimer's disease International, "World alzheimer report 2015: The global impact of dementia - an analysis of prevalence, incidence, cost and trends." http://www.worldalzreport2015.org/downloads/world-alzheimer-report-2015.pdf, 2015. Accessed: 11.06.2017.

[6] World Health Organization & Alzheimer's disease International, "Dementia: A public health priority." http://apps.who.int/iris/bitstream/10665/75263/1/9789241564458_eng.pdf?ua=1, 2012. Accessed: 11.06.2017.

[7] J. Freiherr, J. N. Lundström, U. Habel, and K. Reetz, "Multisensory integration mechanisms during aging," *Frontiers in human neuroscience*, vol. 7, 2013.

[8] L. Shams and A. R. Seitz, "Benefits of multisensory learning," *Trends in cognitive sciences*, vol. 12, no. 11, pp. 411–417, 2008.

[9] S. Shimojo and L. Shams, "Sensory modalities are not separate modalities: plasticity and interactions," *Current opinion in neurobiology*, vol. 11, no. 4, pp. 505–509, 2001.

[10] G. Calvert, C. Spence, and B. E. Stein, *The handbook of multisensory processes*. MIT press, 2004.

[11] A. A. Ghazanfar and C. E. Schroeder, "Is neocortex essentially multisensory?," *Trends in cognitive sciences*, vol. 10, no. 6, pp. 278–285, 2006.

[12] J. Driver and T. Noesselt, "Multisensory interplay reveals crossmodal influences on 'sensory-specific'brain regions, neural responses, and judgments," *Neuron*, vol. 57, no. 1, p. 11, 2008.

[13] P. J. Laurienti, J. H. Burdette, J. A. Maldjian, and M. T. Wallace, "Enhanced multisensory integration in older adults," *Neurobiology of aging*, vol. 27, no. 8, pp. 1155–1163, 2006.

[14] A. M. Peiffer, J. L. Mozolic, C. E. Hugenschmidt, and P. J. Laurienti, “Age-related multisensory enhancement in a simple audiovisual detection task,” *Neuroreport*, vol. 18, no. 10, pp. 1077–1081, 2007.

[15] J. R. Mahoney, P. C. C. Li, M. Oh-Park, J. Verghese, and R. Holtzer, “Multisensory integration across the senses in young and old adults,” *Brain research*, vol. 1426, pp. 43–53, 2011.

[16] K. M. Cienkowski and A. E. Carney, “Auditory-visual speech perception and aging,” *Ear and hearing*, vol. 23, no. 5, pp. 439–449, 2002.

[17] A. Huyse, J. Leybaert, and F. Berthommier, “Effects of aging on audio-visual speech integration,” *The Journal of the Acoustical Society of America*, vol. 136, no. 4, pp. 1918–1931, 2014.

[18] K. Sekiyama, T. Soshi, and S. Sakamoto, “Enhanced audiovisual integration with aging in speech perception: a heightened mcgurk effect in older adults,” *Frontiers in Psychology*, 2014.

[19] A. Setti, K. E. Burke, R. Kenny, and F. N. Newell, “Susceptibility to a multisensory speech illusion in older persons is driven by perceptual processes,” *Frontiers in Psychology*, vol. 4, no. 575, 2013.

[20] N. Tye-Murray, B. Spehar, J. Myerson, M. S. Sommers, and S. Hale, “Cross-modal enhancement of speech detection in young and older adults: Does signal content matter?,” *Ear and hearing*, vol. 32, no. 5, p. 650, 2011.

[21] C. Maguinness, A. Setti, K. Burke, R. A. Kenny, and F. N. Newell, “The effect of combined sensory and semantic components on audio-visual speech perception in older adults,” *Frontiers in aging neuroscience*, vol. 3, p. 19, 2011.

[22] M. M. Murray, C. M. Michel, R. G. de Peralta, S. Ortigue, D. Brunet, S. G. Andino, and A. Schnider, “Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging,” *Neuroimage*, vol. 21, no. 1, pp. 125–135, 2004.

[23] S. Lehmann and M. M. Murray, “The role of multisensory memories in unisensory object discrimination,” *Cognitive Brain Research*, vol. 24, no. 2, pp. 326–334, 2005.

[24] Z. D. Moran, P. Bachman, P. Pham, S. H. Cho, T. D. Cannon, and L. Shams, “Multisensory encoding improves auditory recognition,” *Multisensory research*, vol. 26, no. 6, pp. 581–592, 2013.

[25] A. Thelen, D. Talsma, and M. M. Murray, “Single-trial multisensory memories affect later auditory and visual object discrimination,” *Cognition*, vol. 138, pp. 148–160, 2015.

[26] D. Ueno, K. Masumoto, K. Sutani, and S. Iwaki, “Latency of modality-specific reactivation of auditory and visual information during episodic memory retrieval,” *NeuroReport*, vol. 26, no. 6, pp. 303–308, 2015.

[27] J. Heikkilä, K. Alho, H. Hyvönen, and K. Tiippana, “Audiovisual semantic congruency during encoding enhances memory performance,” *Experimental psychology*, 2015.

[28] J. Heikkilä, K. Alho, and K. Tiippana, “Semantically congruent visual stimuli can improve auditory memory,” *Multisensory Research*, 2017.

[29] J. Heikkilä, K. Alho, and K. Tiippana, “Semantic congruency improves recognition memory performance for both audiovisual and visual stimuli,” *Multisensory research*, 2017.

[30] J. Heikkilä and K. Tiippana, “School-aged children can benefit from audiovisual semantic congruency during memory encoding,” *Experimental brain research*, vol. 234, no. 5, pp. 1199–1207, 2016.

[31] M. A. Cohen, T. S. Horowitz, and J. M. Wolfe, “Auditory recognition memory is inferior to visual recognition memory,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 14, pp. 6008–6010, 2009.

[32] L. Luo, T. Hendriks, and F. I. Craik, “Age differences in recollection: Three patterns of enhanced encoding,” *Psychology and aging*, vol. 22, no. 2, p. 269, 2007.

[33] G. E. Smith and M. W. Bondi, *Textbook of clinical neuropsychology*, ch. Normal aging, mild cognitive impairment, and Alzheimer’s disease. Taylor & Francis, 2008.

[34] T. A. Salthouse, “Decomposing age correlations on neuropsychological and cognitive variables,” *Journal of the International Neuropsychological Society*, vol. 15, no. 05, pp. 650–661, 2009.

[35] T. A. Salthouse, “Initiating the formalization of theories of cognitive aging.,” *Psychology and aging*, vol. 3, no. 1, p. 3, 1988.

[36] P. Kuikka, V. Pulliainen, and R. Hänninen, *Kliininen neuropsykologia.* WSOY, 2001.

[37] F. I. Craik and E. Bialystok, “Cognition through the lifespan: mechanisms of change,” *Trends in cognitive sciences*, vol. 10, no. 3, pp. 131–138, 2006.

[38] T. A. Salthouse, “The processing-speed theory of adult age differences in cognition.,” *Psychological review*, vol. 103, no. 3, p. 403, 1996.

[39] T. A. Salthouse, “Aging and measures of processing speed,” *Biological psychology*, vol. 54, no. 1, pp. 35–54, 2000.

[40] T. A. Salthouse, "Selective review of cognitive aging," *Journal of the International Neuropsychological Society: JINS*, vol. 16, no. 5, p. 754, 2010.

[41] J. D. Kester, A. S. Benjamin, A. D. Castel, and F. I. Craik, "Memory in elderly people," *The handbook of memory disorders*, vol. 2, pp. 543–567, 2002.

[42] J. Juntunen, *Kliininen kognitiviinen neurotiede: Aivot ja ajattelu.* Recallmed Oy, 2008. p. 180.

[43] L. Nyberg, L. Bäckman, K. Erngrund, U. Olofsson, and L.-G. Nilsson, "Age differences in episodic memory, semantic memory, and priming: Relationships to demographic, intellectual, and biological factors," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 51, no. 4, pp. P234–P240, 1996.

[44] T. Hedden and J. D. Gabrieli, "Insights into the ageing mind: a view from cognitive neuroscience," *Nature reviews. Neuroscience*, vol. 5, no. 2, p. 87, 2004.

[45] L.-G. Nilsson, "Memory function in normal aging," *Acta Neurologica Scandinavica*, vol. 107, no. s179, pp. 7–13, 2003.

[46] E. Tulving, "Multiple memory systems and consciousness," *Human neurobiology*, vol. 6, no. 2, pp. 67–80, 1987.

[47] L. R. Squire, J. T. Wixted, and R. E. Clark, "Recognition memory and the medial temporal lobe: a new perspective," *Nature reviews. Neuroscience*, vol. 8, no. 11, p. 872, 2007.

[48] L. R. Squire and S. Zola-Morgan, "The medial temporal lobe memory system," *Science*, vol. 253, no. 5026, p. 1380, 1991.

[49] H. Eichenbaum, A. P. Yonelinas, and C. Ranganath, "The medial temporal lobe and recognition memory," *Annu. Rev. Neurosci.*, vol. 30, pp. 123–152, 2007.

[50] J. J. Medina, "The biology of recognition memory," *Psychiatric Times*, vol. 25, no. 7, pp. 13–15, 2008.

[51] L.-G. Nilsson, L. BÄCkman, K. Erngrund, L. Nyberg, R. Adolfsson, G. Bucht, S. Karlsson, M. Widing, and B. Winblad, "The betula prospective cohort study: Memory, health, and aging," *Aging, Neuropsychology, and Cognition*, vol. 4, no. 1, pp. 1–32, 1997.

[52] T. A. Salthouse, "The aging of working memory.," *Neuropsychology*, vol. 8, no. 4, p. 535, 1994.

[53] F. I. Craik, "Memory changes in normal aging," *Current directions in psychological science*, vol. 3, no. 5, pp. 155–158, 1994.

[54] X. Liu and D. Yan, "Ageing and hearing loss," *The Journal of pathology*, vol. 211, no. 2, pp. 188–197, 2007.

[55] J. E. Peelle and A. Wingfield, "The neural consequences of age-related hearing loss," *Trends in neurosciences*, vol. 39, no. 7, pp. 486–497, 2016.

[56] D. B. Elliot, K. C. Yang, and D. Whitaker, "Visual acuity changes throughout adulthood in normal, healthy eyes: seeing beyond 6/6.," *Optometry & Vision Science*, vol. 72, no. 3, pp. 186–191, 1995.

[57] C. Owsley, "Aging and vision," *Vision research*, vol. 51, no. 13, pp. 1610–1622, 2011.

[58] G. J. Tortora and B. H. Derrickson, *Principles of anatomy and physiology.* John Wiley & Sons, 14 ed., 2014.

[59] P. J. Laurienti and C. E. Hugenschmidt, *Multisensory development*, ch. Multisensory processes in old age, pp. 251–272. Oxford University Press, Oxford, 2012.

[60] M. S. Sommers, N. Tye-Murray, and B. Spehar, "Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults," *Ear and hearing*, vol. 26, no. 3, pp. 263–275, 2005.

[61] N. Tye-Murray, M. Sommers, B. Spehar, J. Myerson, and S. Hale, "Aging, audiovisual integration, and the principle of inverse effectiveness," *Ear and hearing*, vol. 31, no. 5, p. 636, 2010.

[62] A. H. Winneke and N. A. Phillips, "Does audiovisual speech offer a fountain of youth for old ears? an event-related brain potential study of age differences in audiovisual speech perception.," *Psychology and aging*, vol. 26, no. 2, p. 427, 2011.

[63] A. Jesse and E. Janse, "Audiovisual benefit for recognition of speech presented with single-talker noise in older listeners," *Language and cognitive processes*, vol. 27, no. 7-8, pp. 1167–1191, 2012.

[64] H. McGurk and J. MacDonald, "Hearing lips and seeing voices," *Nature*, vol. 264, no. 5588, pp. 746–748, 1976.

[65] M. Alm and D. Behne, "Audio-visual speech experience with age influences perceived audio-visual asynchrony in speech," *The Journal of the Acoustical Society of America*, vol. 134, no. 4, pp. 3001–3010, 2013.

[66] M. Alm and D. Behne, "Age mitigates the correlation between cognitive processing speed and audio-visual asynchrony detection in speech," *The Journal of the Acoustical Society of America*, vol. 136, no. 5, pp. 2816–2826, 2014.

[67] V. Krull and L. E. Humes, "Text as a supplement to speech in young and older adults," *Ear and hearing*, vol. 37, no. 2, pp. 164–176, 2016.

[68] A. Thelen and M. M. Murray, "The efficacy of single-trial multisensory memories," *Multisensory research*, vol. 26, no. 5, pp. 483–502, 2013.

[69] T. R. Schneider, A. K. Engel, and S. Debener, "Multisensory identification of natural objects in a two-way crossmodal priming paradigm," *Experimental psychology*, vol. 55, no. 2, pp. 121–132, 2008.

[70] H. Stanislaw and N. Todorov, "Calculation of signal detection theory measures," *Behavior research methods, instruments, & computers*, vol. 31, no. 1, pp. 137–149, 1999.

[71] N. A. Macmillan, *Signal detection theory as data analysis method and psychological decision model.* Lawrence Erlbaum Associates, Inc, 1993.

[72] N. A. Macmillan and H. L. Kaplan, "Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates.," *Psychological bulletin*, vol. 98, no. 1, p. 185, 1985.

[73] H. Coolican, *Research methods and statistics in psychology.* Psychology Press, 2014.

[74] E. R. Girden, *ANOVA: Repeated measures.* Sage Publications, Inc, 1992.

[75] P. A. Luce and D. B. Pisoni, "Recognizing spoken words: The neighborhood activation model," *Ear and hearing*, vol. 19, no. 1, p. 1, 1998.

[76] X. Z. Liu, L. R. Xu, A. Sismanis, Y. Hu, S. L. Zhang, W. E. Nance, and Y. Xu, "Epidemiological studies on hearing impairment with reference to genetic factors in sichuan, china," *Annals of Otology, Rhinology & Laryngology*, vol. 110, no. 4, pp. 356–363, 2001.

[77] H. M. Bharadwaj, S. Verhulst, L. Shaheen, M. C. Liberman, and B. G. Shinn-Cunningham, "Cochlear neuropathy and the coding of supra-threshold sound," *Frontiers in systems neuroscience*, vol. 8, 2014.

[78] S. Molholm, W. Ritter, D. C. Javitt, and J. J. Foxe, "Multisensory visual–auditory object recognition in humans: a high-density electrical mapping study," *Cerebral Cortex*, vol. 14, no. 4, pp. 452–465, 2004.

[79] W. D. Hairston, M. T. Wallace, J. W. Vaughan, B. E. Stein, J. L. Norris, and J. A. Schirillo, "Visual localization ability influences cross-modal bias," *Journal of cognitive neuroscience*, vol. 15, no. 1, pp. 20–29, 2003.

[80] A. Baddeley, "Working memory: theories, models, and controversies," *Annual review of psychology*, vol. 63, pp. 1–29, 2012.

[81] O. Diaconescu, Andreea, L. Hasher, and A. R. McIntosh, "Visual dominance and multisensory integration changes with age," *Neuroimage*, vol. 65, pp. 152–166, 2013.

[82] M. C. Potter, "Very short-term conceptual memory," *Memory & Cognition*, vol. 21, no. 2, pp. 156–161, 1993.

[83] M. C. Potter, “Conceptual short term memory in perception and thought,” *Frontiers in Psychology*, vol. 3, 2012.

[84] Y.-C. Chen and C. Spence, “When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures,” *Cognition*, vol. 114, no. 3, pp. 389–404, 2010.

[85] P. J. Laurienti, R. A. Kraft, J. A. Maldjian, J. H. Burdette, and M. T. Wallace, “Semantic congruence is a critical factor in multisensory behavioral performance,” *Experimental brain research*, vol. 158, no. 4, pp. 405–414, 2004.

[86] C. Alain and D. L. Woods, “Age-related changes in processing auditory stimuli during visual attention: evidence for deficits in inhibitory control and sensory memory.,” *Psychology and aging*, vol. 14, no. 3, p. 507, 1999.

A Normality tests: CDF graphs and a table

In this appendix, the Cumulative Distribution Functions (CDF) are represented for each dependent variable (d', hit rate, c, reaction time) in each condition (sounds-only, sounds with pictures, sounds with text, spoken words-only, spoken words with pictures, spoken words with text) to graphically show that all data is approximately normally distributed. In the graphs the blue line represents the measured data and the red line the predicted data that is perfectly normally distributed. As it can be noticed, all data (blue line) follow reasonably the predicted data (red line) and thus it can be stated that the normality assumption of sphericity holds.

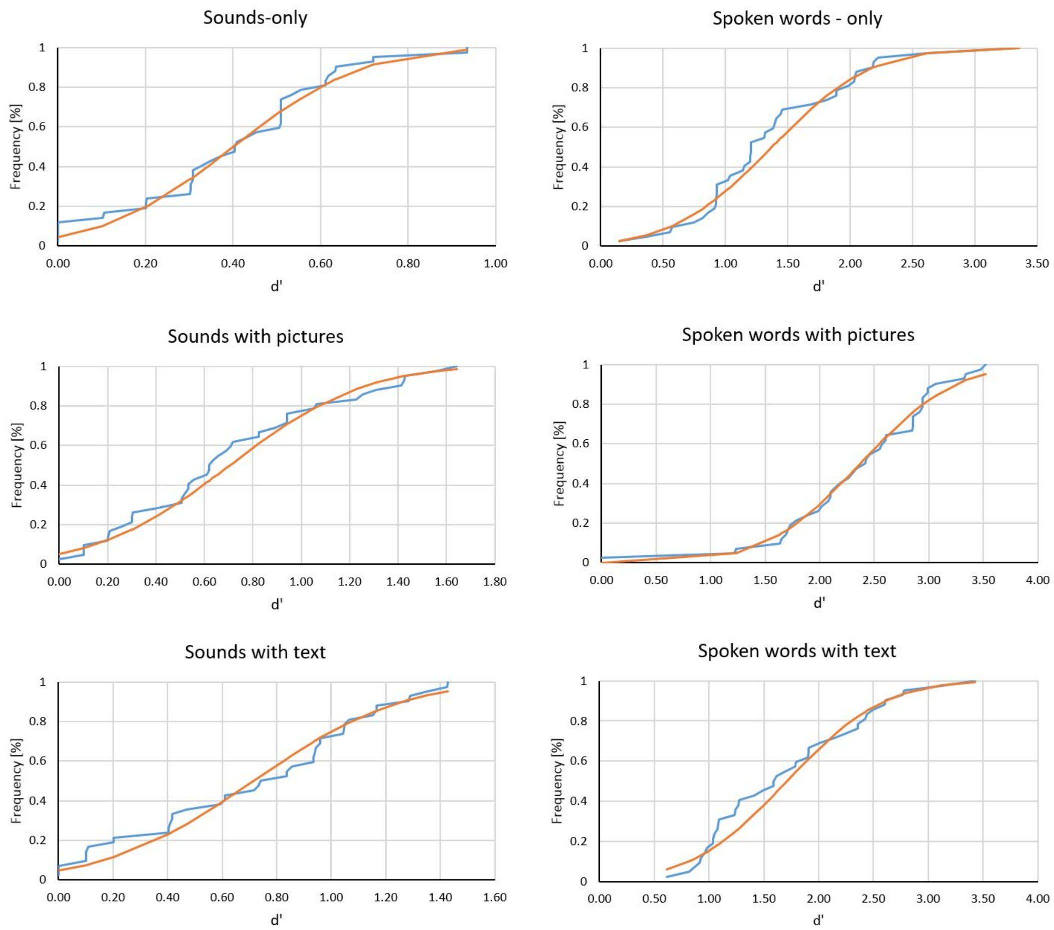


Figure A1: CDF for dependent variable *d'* in each condition: all data approximately normally distributed.

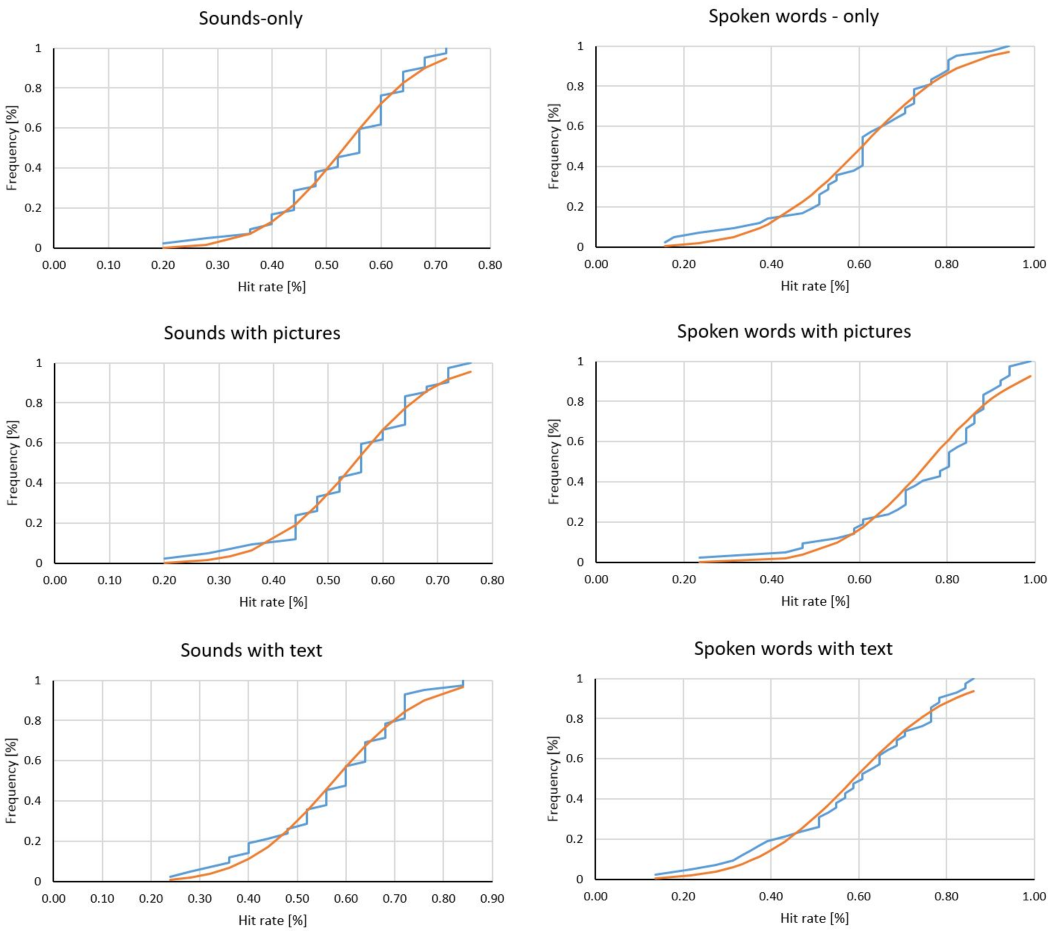


Figure A2: CDF for dependent variable *hit rate* in each condition: all data approximately normally distributed.

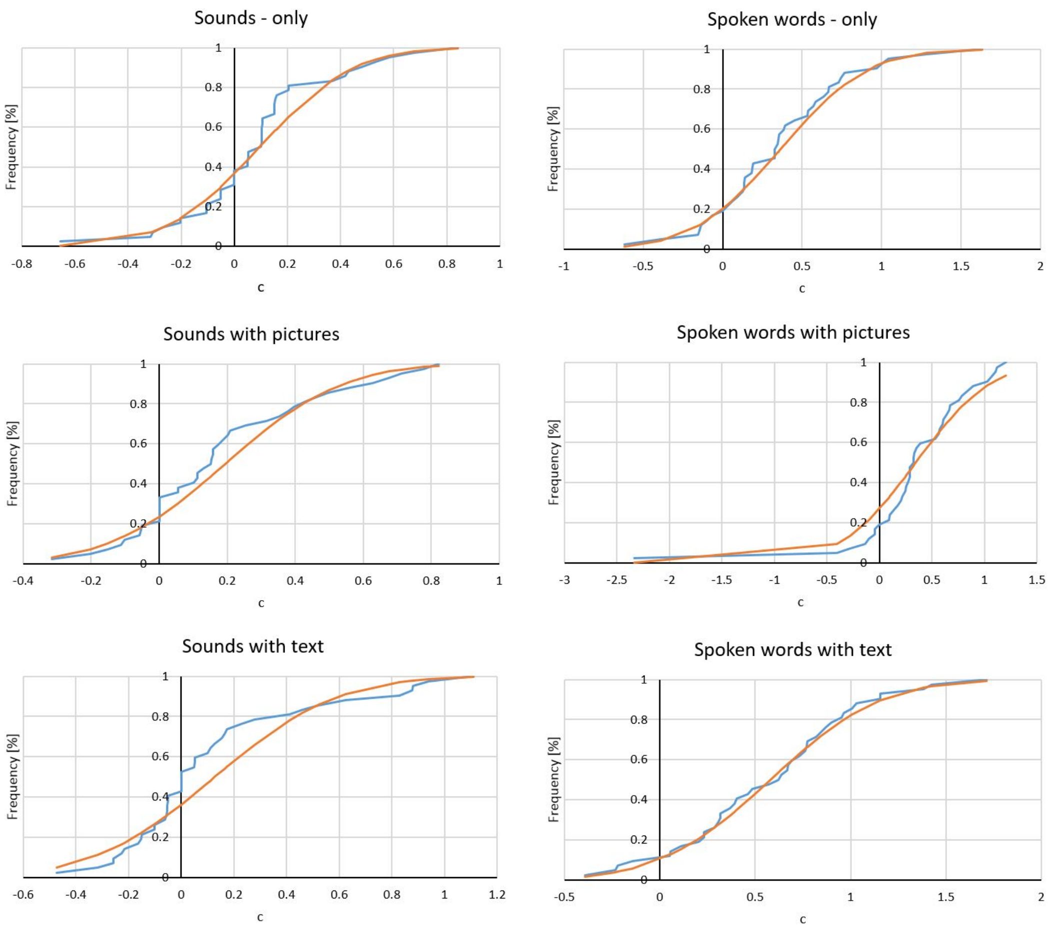


Figure A3: CDF for dependent variable c in each condition: all data approximately normally distributed.

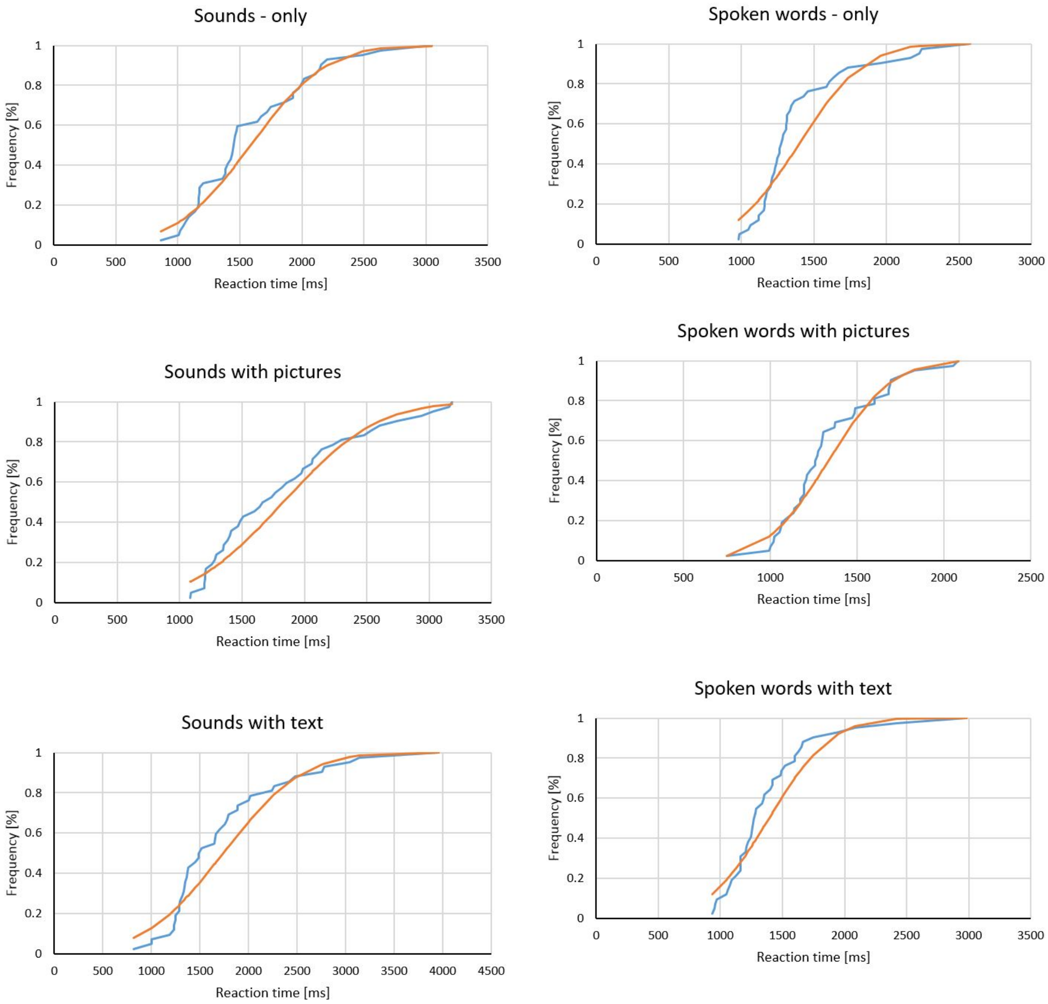


Figure A4: CDF for dependent variable *reaction time* in each condition: all data approximately normally distributed.

Table A1: Double checking normality with a rule of thumb that says that normality can be checked by looking at the differences between the mean and the median; if the difference is less than a half a standard deviation, normality can be assumed (marked as "1").

		Mean	Median	SD	Difference	$\frac{1}{2}SD$	Normality
d'	s	0.402	0.406	0.237	0.005	0.119	1
	sp	0.706	0.627	0.439	0.078	0.220	1
	st	0.715	0.789	0.431	0.074	0.215	1
	w	1.373	1.203	0.635	0.170	0.318	1
	wp	2.371	2.422	0.696	0.051	0.348	1
	wt	1.713	1.604	0.708	0.109	0.354	1
HR	s	0.531	0.560	0.118	0.029	0.059	1
	sp	0.548	0.560	0.125	0.012	0.062	1
	st	0.575	0.600	0.145	0.025	0.073	1
	w	0.606	0.608	0.180	0.001	0.090	1
	wp	0.758	0.804	0.163	0.046	0.081	1
	wt	0.589	0.608	0.181	0.019	0.090	1
c	s	0.096	0.103	0.281	0.007	0.140	1
	sp	0.195	0.153	0.275	0.042	0.137	1
	st	0.129	0.000	0.370	0.129	0.185	1
	w	0.364	0.337	0.442	0.027	0.221	1
	wp	0.347	0.323	0.577	0.025	0.288	1
	wt	0.580	0.629	0.463	0.049	0.232	1
RT	s	1588	1450	487	138	243	1
	sp	1831	1700	601	130	301	1
	st	1739	1502	655	238	328	1
	w	1396	1277	362	119	181	1
	wp	1333	1263	293	70	147	1
	wt	1391	1275	399	116	199	1

B List of stimuli

TEST 1 – Sounds-only

Sounds to be memorized during the study phase (old sounds)

1	Wolf	6	Violin	11	Mixer	16	Sawing	21	Ambulance
2	Dolphin	7	Xylophone	12	Pinball	17	Bike bell	22	Bus horn
3	Seagull	8	Trumpet	13	Roulette	18	Video game	23	Fencing
4	Cicada	9	Kettle	14	Hair spray	19	River	24	Heartbeat
5	Seal	10	Knife	15	Zipping	20	Fax	25	Hiccup

New sounds presented during the test phase together with the old sounds

26	Goose	31	Cymbalo	36	Boiling water	41	Wine pouring	46	Helicopter
27	Rooster	32	Bells	37	Tap dancing	42	Clocks	47	Trolley bell
28	Cuckoo	33	Banjo	38	Cards	43	Typewriter	48	Snore
29	Chick	34	Toaster	39	Brooming	44	Thunder	49	Baby cry
30	Pigeon	35	Hammer	40	Writing	45	White noise	50	Sigh

TEST 2 – Sounds with pictures

Sounds to be memorized during the study phase (old sounds)

1	Lion	6	Horse	11	Washing machine	16	Whistle	21	Chain
2	Cat	7	Guitar	12	Drill	17	Bell	22	Train
3	Duck	8	Accordion	13	Bowling	18	Phone	23	Harp
4	Fly	9	Flute	14	Table tennis	19	Waterfall	24	High heels
5	Turtle	10	Keys	15	Door	20	Ice cubes	25	Printer

New sounds presented during the test phase together with the old sounds

26	Tiger	31	Sheep	36	Can (opening)	41	Toilet flush	46	Cannon
27	Donkey	32	Bongo	37	Sawmill	42	Cuckoo Clocks	47	Car starting
28	Bird	33	Triangle	38	Race car	43	Boat horn	48	Castanets
29	Crocodile	34	Tambourine	39	Karate	44	Wind	49	Yawn
30	Guinea pig	35	Scissors	40	Wine opening	45	Fizzy	50	Keyboard

TEST 3 – Sounds with written words

Sounds to be memorized with simultaneously presented text during the study phase (old sounds)

	SOUND	TEXT (in FIN)		SOUND	TEXT (in FIN)
1	Laugh	Nauru	14	Golf	Golf
2	Bear	Karhu	15	Fire	Tuli
3	Dog	Koira	16	Lock	Lukko
4	Pig	Sika	17	Laser	Laser
5	Goat	Vuohi	18	Book	Kirja
6	Tern	Tiira	19	Stapler	Nitoja
7	Piano	Piano	20	Cough	Yskä
8	Vacuum cleaner	Imuri	21	Rain	Sade
9	Horn	Torvi	22	Coins	Lantit
10	Dice	Noppa	23	Movie roll	Kela
11	Glass	Lasi	24	Car brakes	Jarru
12	Tennis	Tennis	25	Owl	Pöllö
13	Arrow	Nuoli			

New sounds presented during the test phase together with the old sounds

26	Sneeze	31	Chimpanzee	36	Lighter	41	Knocking	46	Biting
27	Elephant	32	Music box	37	Claps	42	Hair dryer	47	Sink
28	Bee	33	Microwave oven	38	Razor	43	Match	48	Cow bell
29	Frog	34	Bagpipes	39	Baseball	44	Fire alarm	49	Scooter
30	Cow	35	Coffee maker	40	Tooth brushing	45	Plane	50	Whale

TEST 4 – Spoken words -only

Spoken words to be memorized during the study phase (old words)

	WORD (in FIN)	TRANSLATION		WORD (in FIN)	TRANSLATION
1	Amme	Bathtub	27	Liitu	Chalk
2	Aski	Box	28	Lumme	Water lily
3	Harppi	Compass	29	Myssy	Woolly hat
4	Aita	Fence	30	Mänty	Pine
5	Hiili	Coal	31	Naali	Arctic fox
6	Huone	Room	32	Mehu	Juice
7	Härkä	Bull	33	Nokka	Beak
8	Jalka	Leg	34	Nukke	Doll
9	Jänis	Hare	35	Oksa	Branch
10	Kaasu	Gas	36	Peili	Mirror
11	Aamu	Morning	37	Pelto	Field
12	Kansi	Lid	38	Ranta	Beach
13	Katto	Roof	39	Rapu	Crab
14	Kaulus	Collar	40	Saari	Island
15	Kelkka	Sledge	41	Setä	Uncle
16	Keppi	Stick	42	Satu	Fairytale
17	Kerho	Club	43	Suola	Salt
18	Kirkko	Church	44	Talvi	Winter
19	Puisto	Park	45	Tuoli	Chair
20	Kori	Basket	46	Tehdas	Factory
21	Koulu	School	47	Tilli	Dill
22	Kuitti	Receipt	48	Tyttö	Girl
23	Kurki	Crane	49	Vilja	Grain
24	Lanka	Thread	50	Viini	Wine
25	Laulu	Song	51	Vaate	Clothing
26	Leipä	Bread			

New spoken words presented during the test phase together with the old words

	WORD (in FIN)	TRANSLATION		WORD (in FIN)	TRANSLATION
52	Häntä	Tail	78	Lumi	Snow
53	Aalto	Wave	79	Lääke	Medicine
54	Kaivos	Mine	80	Maali	Paint
55	Kevät	Spring	81	Maito	Milk
56	Kynä	Pen	82	Metsä	Forest
57	Köysi	Rope	83	Minkki	Mink
58	Linna	Castle	84	Niitty	Field
59	Mato	Worm	85	Otsa	Forehead
60	Nenä	Nose	86	Palkka	Pay
61	Parta	Beard	87	Patja	Matress
62	Pankki	Bank	88	Pihka	Reisis
63	Sauna	Sauna	89	Korppu	Cracker
64	Sohva	Couch	90	Ruoho	Grass
65	Teräs	Steel	91	Salkku	Briefcase
66	Aula	Lobby	92	Seinä	Wall
67	Tulkki	Interpreter	93	Sirppi	Sickle
68	Tassu	Paw	94	Sormi	Finger
69	Vuori	Mountain	95	Syksy	Autumn
70	Kangas	Fabric	96	Tuhka	Ash
71	Kaste	Dew	97	Turkki	Fur
72	Kesä	Summer	98	Varjo	Shadow
73	Koti	Home	99	Varpu	Twig
74	Kulta	Gold	100	Viiva	Line
75	Kuusi	Fir	101	Öljy	Oil
76	Lanttu	Rutabaga	102	Kurssi	Course
77	Leivos	Pastry			

TEST 5 – Spoken words with pictures

Spoken words to be memorized with simultaneously presented pictures during the study phase (old words)

	WORD (in FIN)	TRANSLATION		WORD (in FIN)	TRANSLATION
1	Apila	Clover	27	Kaali	Cabbage
2	Hiiri	Mouse	28	Kakku	Cake
3	Höyhen	Feather	29	Kehto	Cradle
4	Kampa	Comb	30	Kiekko	Puck
5	Tossu	Slipper	31	Kielo	Lily of the valley
6	Kala	Fish	32	Koukku	Hook
7	Mursu	Walrus	33	Kulho	Bowl
8	Herne	Pea	34	Lahja	Present
9	Kuokka	Hoe	35	Leija	Kite
10	Mekko	Dress	36	Lippu	Flag
11	Kaktus	Cactus	37	Luu	Bone
12	Mäyrä	Badger	38	Maissi	Corn
13	Kenkä	Shoe	39	Nappi	Button
14	Sello	Cello	40	Nauris	Turnip
15	Allas	Tub	41	Nuija	Mallet
16	Keinu	Swing	42	Piippu	Pipe
17	Jojo	yoyo	43	Portti	Gate
18	Hylly	Bookshelf	44	Marja	Berry
19	Sormus	Ring	45	Ruuvi	Screw
20	Kurkku	Cucumber	46	Saapas	rubber boot
21	Harja	Brush	47	Saukko	Otter
22	Hattu	Hat	48	Tatti	Boletus
23	Hella	Stove	49	Teltta	Tent
24	Housut	Pants	50	Tussi	Marker pen
25	Joutsen	Swan	51	Varis	Crow
26	Juusto	Cheese			

New spoken words presented during the test phase together with the old words

	WORD (in FIN)	TRANSLATION		WORD (in FIN)	TRANSLATION
52	Aura	Plow	78	Neula	Needle
53	Heinä	Hay	79	Nilkka	Ankle
54	Hiha	Sleeve	80	Paino	Weight
55	Hiiva	Yeast	81	Passi	Passport
56	Jäsen	Member	82	Pata	Pot
57	Jälki	Footprint	83	Pinni	Pin
58	Jono	Queue	84	Pitsa	Pizza
59	Kaarna	Bark	85	Puikko	Stick
60	Kasvi	Plant	86	Pussi	Bag
61	Kasvot	Face	87	Ripsi	Eyelash
62	Kartta	Map	88	Ruutu	Square
63	Keihäs	Spear	89	Saavi	Tub
64	Kenttä	Field	90	Savu	Smoke
65	Ilves	Lynx	91	Soija	Soy
66	Korva	Ear	92	Silta	Bridge
67	Kuppi	Cup	93	Taksi	Taxi
68	Kirje	Letter	94	Taimi	Seedling
69	Kylki	Rib	95	Takki	Coat
70	Laama	Llama	96	Tonttu	Christmas elf
71	Latu	Ski trail	97	Torni	Tower
72	Lasso	Lasso	98	Tuoli	Chair
73	Lehti	Leaf	99	Tukki	Log
74	Liha	Meat	100	Tupsu	Tassel
75	Lyhde	Sheaf	101	Vihta	Bath whisk
76	Metro	Metro	102	Vihko	Notebook
77	Riekko	Willow grouse			

TEST 6 – Spoken words with written words

Spoken words to be memorized with simultaneously presented written words during the study phase (old words)

	WORD (in FIN)	TRANSLATION		WORD (in FIN)	TRANSLATION
1	Hana	Tap	27	Huivi	Scarf
2	Jyvä	Grain	28	Täti	Aunt
3	Järvi	Lake	29	Lauta	Board
4	Kanto	Stump	30	Päivä	Day
5	Koru	Jewellery	31	Kaula	Neck
6	Laakso	Valley	32	Vatsa	Belly
7	Leima	Stamp	33	Piimä	Sour milk
8	Myyrä	Mole	34	Kuja	Alley
9	Nasta	Thumb pin	35	Myyjä	Salesman
10	Peitto	Blanket	36	Akku	Battery
11	Pilvi	Cloud	37	Arpi	Scar
12	Ryhmä	Group	38	Hame	Skirt
13	Seppä	Blacksmith	39	Hauki	Pike
14	Tammi	Oak	40	Hillo	Jam
15	Taulu	Painting	41	Kaappi	Closet
16	Varas	Thief	42	Katu	Street
17	Verho	Curtain	43	Kettu	Fox
18	Kauppa	Shop	44	Koivu	Birch
19	Poro	Reindeer	45	Lakka	Lacquer
20	Runko	Trunk of a tree	46	Luola	Cave
21	Tassu	Paw	47	Matto	Rug
22	Rotta	Rat	48	Niemi	Cape
23	Valo	Light	49	Piste	Point
24	Puuro	Porridge	50	Serkku	Cousin
25	Sänky	Bed	51	Verkko	Net
26	Äiti	Mother			

New spoken words presented during the test phase together with the old words

	WORD (in FIN)	TRANSLATION
52	Ansa	Trap
53	Haapa	Aspen
54	Hiekka	Sand
55	Hissi	Elevator
56	Ilta	Evening
57	Kahvi	Coffee
58	Keitto	Soup
59	Kerma	Cream
60	Kieli	Tongue
61	Kivi	Stone
62	Käpy	Cone
63	Leuka	Chin
64	Luomi	Eye lid
65	Malli	Model
66	Masto	Mast
67	Muuri	Brick wall
68	Mökki	Cottage
69	Neula	Needle
70	Nivel	Joint
71	Pahvi	Cardboard
72	Paita	Shirt
73	Pappi	Preast
74	Pihvi	Steak
75	Porras	Step
76	Posti	Mail
77	Pyyhe	Towel
78	Ratas	Cogwheel
79	Riisi	Rice
80	Ruusu	Rose
81	Sammal	Moss
82	Sieni	Mushroom
83	Silmä	Eye
84	Solmu	Knot
85	Sukka	Sock
86	Sumu	Fog
87	Sähkö	Electricity
88	Talo	House
89	Tanssi	Dance
90	Tikka	Woodpecker
91	Tyyny	Pillow
92	Tähti	Star
93	Varvas	Toe
94	Viiksi	Moustache
95	Pensas	Bush
96	Poski	Cheek
97	Mauste	Spice
98	Mummo	Granny
99	Loska	Slush
100	Puuma	Puma
101	Pulla	Bun
102	Vaahto	Foam

C Participant consent forms

Tiedote ja suostumuslomake

Tutkimuksessa selvitetään tunnistusmuistin toimintaa ja aistitiedon käsittelyä 65 vuotta täyttäneillä henkilöillä. Tutkimus on perustutkimusta, josta ei ole suoranaista hyötyä tutkittavalle eivätkä tulokset kerro muistin toiminnasta käytännössä.

Kokeessa esitetään tavallisia ääniä tai puhetta kuulokkeilla ja kuvia tai tekstiä tietokoneen näytöllä (esim. tuulen ääni, kukan kuva, sana "kissa" puhuttuna tai kirjoitettuna). Näihin reagoidaan painamalla näppäimistön painikkeita ohjeiden mukaan. Koe sisältää muistiin painamisosan ja muistista palautusosan. Se kestää noin tunnin.

Osallistuminen on täysin vapaaehtoista ja voit halutessasi koska tahansa keskeyttää tai lopettaa kokeen seuraamuksitta ja syytä ilmoittamatta. Tulokset käsitellään luottamuksellisesti ja nimettömästi eli nimeäsi ei yhdistetä mittaustuloksiin tai muihin tietoihisi. Tulokset raportoidaan ryhmätasolla tieteellisessä raportissa. Aineisto arkistoidaan lukittuihin tiloihin Helsingin yliopistolle ja anonyymi sähköinen aineisto salasanan taakse.

Lisätietoja saa tutkimuksen suorittavalta opiskelijalta Petra Fagerlundilta (puh. 0503567081, Aalto-yliopisto) ja hänen ohjaajaltaan Kaisa Tiippanalta (kaisa.tiippana@helsinki.fi, puh. 0503185721, vastuullinen tutkija, Helsingin yliopisto). Tutkimukselle ei ole erillistä rahoitusta, vaan se on osa yliopistojen perustutkimusta ja tutkinto-opetusta.

Allekirjoitukseni ilmaisee, että olen lukenut ja ymmärtänyt tämän tekstin ja että suostun osallistumaan tutkimukseen.

Allekirjoitus: ____

Nimenselvennys: ____

______________________________, _____ . _____ .2016

paikkakunta päivämäärä

Koehenkilön taustatietolomake

koehenkilötunnus:_________

Ikä: _________ vuotta

Ympäröi oikea vaihtoehto ja kirjoita tarvittaessa lisätietoja:

Sukupuoli: nainen / mies / en osaa sanoa

Äidinkieli: suomi / muu

Onko näössä tai kuulossa havaittu mitään poikkeavaa?

Ei / Kyllä

Onko todettu oppimis- tai muistiongelmia tai neurologisia diagnooseja?

Ei / Kyllä